



---

German  
Historical  
Institute  
London

---

# German Historical Institute London Bulletin

*Conference Report: Contemporary Historians and the Reuse of Social  
Science-Generated Datasets: An International Dialogue on the Challenges  
Presented by Social Science Data*

by Clemens Villinger

*German Historical Institute London Bulletin*  
Vol. XLIV, No. 1 (May 2022), 150–6

ISSN 0269-8552

*Contemporary Historians and the Reuse of Social Science-Generated Datasets: An International Dialogue on the Challenges Presented by Social Science Data.* Workshop organized by the German Research Foundation (DFG) project ‘Social Science Data as Sources for Contemporary History’ (‘Sozialdaten als Quellen der Zeitgeschichte’) and held at the German Historical Institute London on 28–30 October 2021. Conveners: Lutz Raphael (Trier University), Sabine Reh (Research Library for the History of Education, BBF-DIPF Berlin), Pascal Siegers (GESIS—Leibniz Institute for the Social Sciences), Kerstin Brückweh (Berliner Hochschule für Technik), and Christina von Hodenberg (GHIL).

The third workshop of the DFG-funded project ‘Social Science Data as Sources for Contemporary History’ aimed to establish an international dialogue between historians, sociologists, and representatives of the infrastructure that collects and provides access to social science-generated datasets. In her introductory remarks, Christina von Hodenberg (GHIL) emphasized the value of international exchange for a reflection of the different approaches employed by contemporary historians who analyse and incorporate social science data into their research. In addition to addressing such methodological questions, the workshop provided an opportunity to gain a deeper understanding of how research data infrastructures process, archive, and make social science data accessible in countries across Europe.

The first panel focused on the reuse of qualitative and life history interviews. In her presentation, the sociologist Jane Gray (Maynooth University) introduced her research on ‘family rhythms’, reusing and combining archived qualitative social science data from the Life Histories and Social Change Collection, and drawing on semi-structured interviews conducted during the national longitudinal study of children Growing Up in Ireland (GUI). Both datasets have been deposited in the Irish Qualitative Data Archive and are now maintained and disseminated by the Digital Repository of Ireland. According to Gray, working across datasets using descriptive approaches and mixing them with other historical sources such as quantitative data allows an analysis of changing relationships between children and their grandparents across extended periods of time. Gray discussed the

implications of a 'descriptive turn' in the social sciences, which draws on diverse sources (such as qualitative records and social media data) to present complex social phenomena and social change, thus overcoming the limitations of traditional surveys.

Clemens Villinger (GESIS) explained how he reused interviews conducted by social scientists in East Germany during the 1990s to write a history of consumption from an everyday perspective. He identified three major obstacles to his research: first, the interviews were hard to locate because they remained in the personal archives of the interviewers; second, access depended on personal sympathies; and third, different ethical views exist about whether they can be reused at all. He called for a code of ethics to make it easier for historians to reuse social science data responsibly, and to reduce the associated costs. Taking his own research on the attribution of consumer responsibility after 1989–90 as an example, Villinger argued that the benefits of reusing interviews outweigh the challenges, while support was missing to reduce the burden of analysing existing social science research data.

The third paper was presented by Mary Stewart and Charlie Morgan (both British Library Sound Archive). Unlike in Germany, the reuse of interviews has played a key role in the British oral history movement since its beginnings in the 1970s. This is why the Sound Archive aims to make as much qualitative data as possible available not only to scientists and academics, but also to the media, artists, and families. Before the public reuse of older collections is permitted, the General Data Protection Regulation (GDPR) requires the archive to identify personal data about living people and to evaluate whether its public release is likely to cause 'substantial damage and distress'. The nature of the interviews requires an elaborate lexicon search engine to identify sensitive passages. This can itself be reused to make the collection searchable and therefore more accessible. Unlike in Germany, the British Library Sound Archive datasets are not anonymized, which means that information is not lost when they are used for research.

In her comment, Kerstin Brückweh (BHT) suggested establishing a help desk for historians dealing with ethical questions. She also raised the question of whose history we are writing if interviews

with 'ordinary people' are less easily accessible than interviews with 'movers and shakers' such as politicians.

The second panel focused on survey data as a source for social history and started with a presentation by Mor Geller (Hebrew University of Jerusalem) about the KINO-DDR social science research project carried out by the East German Central Institute for Youth Research. The project was designed to elicit viewers' opinions of socialist films. From a history of knowledge perspective, Geller demonstrated the complex relations between the social scientists, the survey, and the participants, which she characterized as 'a double-ended line of communication'. She argued that opinion polls can be used as a historical source to open up ways of studying the relationship between the socialist state and its population.

Marcus Böick (Ruhr-University Bochum) spoke about his analysis of interviews with managers working for the Treuhandanstalt (Trust Agency, set up in 1990 to privatize East German enterprises), which were conducted by an ethnologist in 1992. After tracking them down in the personal possession of a former employee, Böick managed to retrieve the interview transcripts from a number of floppy discs. He used them to write a social microhistory of the Treuhandanstalt from the perspective and experience of the mostly West German managers. By identifying narrative patterns, Böick was able to create types such as the 'industry manager' and topoi such as the self-definition as pioneers working on the economic frontier of the 'wild east'. Böick highlighted open questions concerning the use of datasets rediscovered by historians when there are no guidelines for their appropriate use.

Moritz J. Feichtinger (University of Bern) then introduced his work on quantification practices used to monitor, model, and manipulate societies. He drew upon the Hamlet Evaluation System (HES) used during the Vietnam War as an example. To understand and analyse computing techniques dating from the 1960s and 1970s, Feichtinger engages with a process he calls data 're-enactment', which consists of five steps: the conversion of data into a readable format; the creation of a data life cycle model; the annotation of converted datasets; the simulation (or mimicking) of historic update, maintenance, aggregation, and query routines; and finally, publication as a web-based

simulation. According to Feichtinger, this approach allows a deeper understanding of how the use of data shaped (military) representations of the world that not only influenced decision-making and policy-making processes, but also had a tangible impact on the Vietnamese people.

The comment was given by Christina von Hodenberg. She asked what theoretical, ethical, and practical aspects need to be considered when reusing social science data produced in dictatorships, wars, or colonial contexts. During the discussion both Böick and Pascal Siegers emphasized that errors, biases, and self-censorship are typical of data production in all political contexts. The second part of the discussion revolved around the fundamental question of whether it makes sense to reuse social science data if they do not allow established historical narratives to be challenged.

The second day started with a presentation by Irena Saleniece (Daugavpils University) on oral history interviews with Latvian teachers that are archived in the Centre of Oral History established in 2003. Saleniece is conducting new interviews and reusing existing qualitative datasets to write an experiential history of the Sovietization of the Latvian school system between the 1940s and 1960s. For her, oral history interviews with different generations of (often bi- or trilingual) Latvians serve to counteract the record from the state archives, which during the Soviet period falsified facts and silenced inconvenient voices. She focuses on emotional, episodic, and bottom-up perspectives to break through the standardized 'Bolshevik speak'.

The director of the Mass Observation Archive, Fiona Courage (University of Sussex), gave an introduction to the history and holdings of the archive, as well as her own research on the value of higher education. The initial mass observation project ran from 1937 to the 1950s and was revived in 1981. To this day, the charity-based archive records everyday life in Britain using volunteer panel writers who fill in questionnaires three times a year and also keep diaries. Like the interviews in the British Library Sound Archive, the data are not anonymized. As Courage put it, the broad consent of study participants allows personal data to be used to reconstruct long-term life stories.

In his comment, Pascal Siegers stressed the value of historical research on socialization in schools and other institutions, arguing that historians could enrich the debate in the social sciences. He questioned the reliability of oral history sources, pointing to their subjectivity. In response, Lutz Raphael remarked that oral history interviews could help to reconstruct processes of subjectivization.

The final panel started with a presentation by Alexander Nütznadel (HU Berlin) about the impact of the 'behavioural turn' on economic history. He used examples from the DFG-funded programme 'Experience and Expectation' to explain how reused social science-generated datasets from large-scale surveys can be combined with techniques like 'distant reading' of traditional sources such as newspapers to investigate how interactions between individual preferences, beliefs, and economic expectations led to economic decisions. This historicization of expectations not only raises methodological questions, but also leads to practical problems related to the long-term storage and accessibility of the research data produced. To manage and store the data, the programme has partnered with the Berlin State Library to design an infrastructure based on MyCoRe, which is free, open-source software for the development of data repositories.

The central question of the joint presentation by Benoît Majerus and Lars Wieneke (both Luxembourg Centre for Contemporary and Digital History, C<sup>2</sup>DH) was how clandestine global and local networks of tax evasion can be identified by methods of data extraction from the public register of companies in Luxembourg. The main goal of their project is to identify and analyse networks of individual actors who registered companies. Although the registry is available in standardized PDF documents, named-entity extraction and a data-based understanding of these networks pose complex methodological and technical questions. The data will permit an understanding of how networks for tax evasion developed in Luxembourg from the beginning of the twentieth century.

Lastly, Michael Whittall (Friedrich-Alexander University Erlangen-Nuremberg) outlined a sociological project that revisits interviews with East German works councils conducted in the early 1990s. These historical interviews will be compared with recent sources on works councils in selected companies that are still in existence. Whittall and

his colleagues aim to reconstruct changing perceptions of works councils in relation to factors such as qualifications or length of service. Like all projects represented in the workshop that reuse qualitative data produced by research on the transformation of the 1990s, this project faces data challenges such as accessibility, ethical and ownership questions, difficulties in researching historical production contexts, and issues of long-term storage.

According to Lutz Raphael, who commented on the last panel, all presentations illustrated that the old division of labour between sociology and history is becoming obsolete, not only because of new sources, but also because of changing research methods. The search for weighted factors of causality is increasingly giving way to the search for patterns, meaning, process, and agency. Even though the presentations touched on different subjects and sources, Raphael proposed the category of historical experience as a unifying point that could connect different branches of research. At the same time, he pointed critically to the emergence of a methodological gap produced by computing processes that are no longer fully understood by (most) historians. In response, Andreas Fickers (Luxembourg Centre for Contemporary and Digital History, C<sup>2</sup>DH) described digital hermeneutics as a common space where data, tools, and infrastructure are shared. In his view, historians are now experiencing nothing less than a turning point in the history of science that is fundamentally changing epistemic traditions.

In his concluding comment, Fickers suggested four different modes of reusing social science-generated data: re(dis)covery, reinterpretation, recontextualization, and re-enactment. The first aspect involves historians applying techniques such as retrodigitization, the annotation of metadata, and restoring data that used to be the typical domain of archives or libraries. Reinterpreting data means using new digital tools that not only empower historians, but also limit historical knowledge production. Reflecting on the opportunities and limits of digital methods, Fickers pointed to 'tool criticism' as a new historical instrument that can help to narrow the methodological gap. He argued that recontextualizing data also poses ethical questions that can include disfiguring meaning, while indexation processes can also have excluding effects. To deal with issues arising from reframing sources in a digital

environment, he suggested engaging in practices of ‘ethical editing’ and interface criticism to understand how datasets are (re)presented on digital platforms. His last point on re-enacting referred to the materiality of datasets and the knowledge that is embedded in both the physical datasets and the machines processing them.

The final discussion showed that there was no common understanding of how the terms ‘use’ and ‘reuse’ should be differentiated. But there was agreement that social science datasets are valuable sources that must be secured, archived, and made accessible. Von Hodenberg pointed out that there is a lack of international data infrastructure, even though scientific knowledge production is increasingly dominated by international co-operation. Siegers explained this in terms of the specialization of nation-based scientific communities who demand infrastructure which fits their needs. Fickers, on the other hand, pointed to international standards, such as the Europeana metadata scheme, that not only enable interoperability, but also make archived datasets findable. In the end, the workshop showed that the reuse of datasets by contemporary historians is a dynamic field characterized by decentralized infrastructure and a broad variety of sources, tools, and approaches. It became clear that the collection, organization, and interpretation of social science-generated datasets will continue to be a task for years to come.

CLEMENS VILLINGER (GESIS – Leibniz Institute for the Social Sciences)