

Brian Kleiner, Isabelle Renschler, Boris Wernli, Peter Farago, & Dominique Joye (eds) (2013) *Understanding Research Infrastructures in the Social Sciences*. Zurich: Seismo.

Dominique Vinck

dominique.vinck@unil.ch

Understanding Research Infrastructures in the Social Sciences introduces us to major quantitative data infrastructures developed in the social sciences (large-scale surveys, monitoring of cohorts, big international databases, data access and exchange platforms) and their activities (documentation, data harmonisation and dissemination, quality management, methodological innovations, training and maintenance). Indeed, how can we understand scientific work without taking into account the infrastructures that make it possible? This question has gradually permeated sciences studies, especially *Big Science* where sophisticated instrumentation is used to produce original data: particle accelerators in high energy physics, radio telescopes in astronomy, magnetic resonance scanners in the neurosciences, but also more conventional infrastructures such as herbaria in botany, observation stations in the Antarctic, etc. These material infrastructures greatly depend on organisational (planning and running of expeditions, launch of observation satellites) and coordination work (25 years to design and develop the LHD collider), logistic and maintenance work.

Although these infrastructures often have a very visible side owing to the gigantic nature of their instruments, buildings or collections, they also have a hidden side, notably made up of categories structuring the distribution of objects and activities: nomenclatures in chemistry, classification of conditions in psychiatry, protocols in biomedical research, etc. This invisibility stems not

only from their apparent immateriality but also from the need for a certain degree of efficiency; researchers do not wish to be encumbered with problems relating to filing, logistics, procedures or maintenance. They desire immediate access to the topic of their research or the data they are seeking, put differently, to meet with success, a good infrastructure has to be invisible, together with the equipment making it up and the operations on which it depends, not to mention the foot soldiers carrying out these operations.

The book by Kleiner et al. (2013), is disturbing from this point of view since it precisely strives to make these infrastructures, which researchers and policy-makers alike would prefer not to see, visible. Who really wants to know about the research databases in the social sciences, the protocols and logistics underlying major surveys, or the budgets that need to be allocated to these or the associated institutions and skills that should be stabilised? Even within the social sciences, these infrastructures suffer from a lack of recognition. According to the authors of the book's first chapter, the researchers in our disciplines bestow even less recognition on these infrastructures than those in the natural sciences with respect to their instrumentation. The book targets social science researchers and policy makers concerned with the production of knowledge in order to increase awareness of the importance of these research infrastructures.

The book begins with two background articles. The first, by Renschler, Kleiner and Wernli, proposes a series of concepts to characterise these research infrastructures (RI). By extension, they cover non-specific technical infrastructures (telecommunication networks, IT systems, communication protocols, etc., as well as the associated technical personnel) important for research, and research institutions. However, the authors have opted for a restrictive definition centred on the supply of resources for research (quantitative data, methods and organisations, skills and instruments set up for collecting and processing data). They do not therefore look at the collections of audiovisual, iconic or textual media that are attracting the attention of researchers in the human and social sciences. They raise policy makers' awareness of the importance of the decisions and investments needed to maintain and develop the material and institutional aspects of these RIs. These infrastructures are embedded in and depend on research institutions, communities and practices and regulatory (e.g. concerning the respect of individual freedoms and the protection of privacy), administrative or technical non-specific infrastructures. The setting up of European Research Infrastructure Consortia (ERIC), the Consortium of European Social Science Data Archives (CESSDA) or the International Federation of Data Organisations (IFDO) lends them a new institutional anchoring. They evolve in relation to the needs of researchers and policy makers, as well as technological innovations, while conversely they influence research dynamics, creativity, quality standards and questions on the political agenda.

The second chapter, by Max Kaase, goes back over the genesis of and evolutions undergone by these RIs, since private data archives set up for the social sciences (on public opinion) in the United States in 1946, through to the establishment of national statistical services and the structuring of wide-scale and repeated international surveys (Eurobarometer, European Social Survey). Comparative research has given an impetus to methodological thinking (harmonisation of protocols, quality management, comparability, management of errors, etc.) and to a roadmap for European infrastructures and the setting up of e-infrastructures (remote access to data).

The book then presents eight RI case studies for the social sciences. These are documented in terms of history, end goals, concerned actors, practices and remaining challenges. Generally carried out by authors involved in these RIs, they are preceded by the cross-reading of Wernli, Renschler, Kleiner and Joye who highlights four main components of the work carried out by RIs in the field of social sciences: 1/ *Data documentation, curation, preservation, provision and dissemination* for which the cooperation between researchers and national statistical institutes has been given a new lease of life with the open data movement and the increasing interest in the reuse of qualitative data; 2/ *Data collection* (formulating questions, sampling, data collection procedures, setting up of panels and time series), *harmonisation and interlinking* in order to make international and time-based comparisons possible; 3/ *Improvement of methods and methodological innovation*: quality control and continuous improvement approach, methodological thinking, documentation of the data life cycle (*Data Documentation Initiative - DDI*); 4/ *Teaching best methodological standards and training*, in order to maintain an international pool of skilled people.

Three chapters deal with efforts to coordinate and harmonise databases. Hans Jørgen Marker reports on the work of the CESSDA in terms of strengthening cooperation between European databases. Roxane Silberman discusses the European *Data without Boundary* network regarding the open access to State data: de-identification, access conditions (e.g. to tax or health data), remote and secure access (for researchers but also for firms), obstacles to sharing data, accreditation procedures. Louise Corti explores the British *Qualidata* case of an archive catalogue (notably for interview transcriptions) and its new challenges (archiving online surveys, blogs and micro-messages (notably tweets), data associated with publications (*enhanced publication*) and interoperability).

The ensuing five case studies relate to the running of major surveys. Dean Lillard addresses the international harmonisation of longitudinal data from national panels (*Cross-National Equivalent File - CNEF*). Janet Gornick discusses the *Luxembourg Income Study* (LIS). Rory Fitzgerald

presents the *European Social Survey* (ESS), its operation and methodological evolutions. Paul de Graaf and Loek Halman overview the changes operated on the *European Value Survey* (EVS) focusing on opinions, attitudes and beliefs. Axel Börsch-Supan reports on the *Survey of Health, Ageing and Retirement in Europe* (SHARE). These authors describe RIs from their point of view as stakeholders supporting RIs and facing the instability of national financing. They write pleas in favour of the RIs rather than an analysis of their shaping and results. From this point of view, the book does not stand as a social study of RIs but the quality of the documentation provides fuel for further questions.

Three chapters then follow focusing on methodological innovation and then two on training. Willem Saris presents progress in terms of predicting the quality of the questions used in questionnaires. Ineke Stoop deals with quality management, protection against fraud, data documentation, transparency, metadata about survey implementation (paradata) and assessment of non-response bias. Joachim Wackerow discusses the establishment of a standard for metadata (DDI) in the social sciences. The two chapters on training report on expectations for research design and data analysis (Silke Schneider, Alexia Katsanidou, Laurence Horton and Christof Wolf) and on high-level seminars on methodology (QMSS – *Quantitative Methods in the Social Sciences*) (Angela Dale).

Finally, two chapters outline a number of lessons to be learnt from the experience of RIs in the social sciences. Markus Zücher looks at the institutional and political anchoring of RIs (dependence on project-based financing, institutional division of responsibilities). Kleiner, Renschler, Wernli and Farago conclude the book with an appraisal of what RIs do to the social

sciences: opening up new possibilities, offering access to more data (e.g. for comparative work), improving quality and research efficiency, and converging research practices. They also outline the upcoming challenges stemming from the increasing difficulty of accessing potential survey participants (growing lack of interest in surveys, legal requirements), but also in terms of articulation with other types of data (geolocation, biological markers, digitally native data, qualitative data) and reuse.

This book might be thought of as a contribution to *infrastructure studies*. While it describes a series of RIs, it does not document concrete practices and assembling as its title (*Understanding...*) might suggest. It is more generally speaking a book drafted by the designers and promoters of RIs busy defending the cause of these structures. Both social science researchers and policy makers have ignored this cause, which is why a book targeting these very people had to be put together. However, for science studies researchers keen to reflect on the practices and dynamics of research, the book leaves something to be desired. The authors, who have the skills to question complex social phenomena, seem to have forgotten to use these skills when exploring their own work practices. The book is then a first step towards greater reflexivity within the social science community. Furthermore, whether it explores some challenges (political, institutional and financing-related instability; methodological and technological challenges; threats to human resources), it does not mention big data, while policy makers are beginning to be seduced by emerging actors (IT specialists, big data consultants) and arguments extolling the merits of new data science undermining the worth of costly major surveys.