



German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD

Working Paper Series

Working Paper

No. 50

Developing the Research Infrastructure
in the Social Sciences:
The Role and Contribution of
the German Research Foundation

Eckard Kaemper, Manfred Niessen

December 2008

Working Paper Series of the Council for Social and Economic Data (RatSWD)

The *RatSWD Working Papers* series was launched at the end of 2007. Since 2009, the series has been publishing exclusively conceptual and historical works dealing with the organization of the German statistical infrastructure and research infrastructure in the social, behavioral, and economic sciences. Papers that have appeared in the series deal primarily with the organization of Germany's official statistical system, government agency research, and academic research infrastructure, as well as directly with the work of the RatSWD. Papers addressing the aforementioned topics in other countries as well as supranational aspects are particularly welcome.

RatSWD Working Papers are non-exclusive, which means that there is nothing to prevent you from publishing your work in another venue as well: all papers can and should also appear in professionally, institutionally, and locally specialized journals. The *RatSWD Working Papers* are not available in bookstores but can be ordered online through the RatSWD.

In order to make the series more accessible to readers not fluent in German, the English section of the *RatSWD Working Papers* website presents only those papers published in English, while the German section lists the complete contents of all issues in the series in chronological order.

Starting in 2009, some of the empirical research papers that originally appeared in the *RatSWD Working Papers* series will be published in the series *RatSWD Research Notes*.

The views expressed in the *RatSWD Working Papers* are exclusively the opinions of their authors and not those of the RatSWD.

The RatSWD Working Paper Series is edited by:

Chair of the RatSWD (2007/ 2008 Heike Solga; 2009 Gert G. Wagner)

Managing Director of the RatSWD (Denis Huschka)

Developing the Research Infrastructure in the Social Sciences: The Role and Contribution of the German Research Foundation

By Eckard Kaemper and Manfred Niessen

DFG (German Research Foundation)

Abstract

The DFG's (German Research Foundation) strategy for future data research infrastructures should be based on what has been achieved thus far and the lessons that can be learned: First, the focus should be on "providing data" rather than on "sharing data." Second, projects whose primary purpose is to provide a common good should seek to build research infrastructure. The DFG has powerful means at its disposal for funding outstanding infrastructure projects. It is up to the scientific community to adapt and utilize these funding instruments. Strategic cooperation is required among all the interested parties in the field: cooperation on thematic priorities within the research community; cooperation on options for funding between the research community and funding institutions; cooperation on the division of labour between the funding institutions (including ministries), both on the national and on the international level. The DFG is prepared to play an active role in this cooperation under the leadership of its elected bodies (Fachkollegien and Senat).

Keywords: Large Scale Studies; Strategic Cooperation

The research infrastructure of the social sciences—and other disciplines—has long held a place on the DFG’s agenda, both in its funding policies and its funding activities. The DFG provided the funding for both ZUMA¹ and SOEP², for example, and nurtured them in their formative years. DFG has also funded activities at ZA³, IZ⁴ and ZPID⁵. All these activities have been about data—about methods and methodologies for collecting and analysing data, and about organisations and structures for preserving data and making them accessible.

Data-related research infrastructure has become a more prominent topic of research policy in recent months and years, nationally as well as internationally. In the general science policy debate, much emphasis has been placed on “sharing data”, often also referring to open access initiatives.

1. “Sharing Data”—A Realistic Approach

The idea of “sharing data” focuses on data produced in research projects that pursue specific hypothesis and generate the data accordingly. That is, the data are generated or collected to answer the specific questions of the project at hand; thus, the data are project-specific.

It is taken for granted that sharing data will increase efficiency and reduce research costs by necessitating replication studies and reducing duplication in data production. However, data sharing is by no means a new idea. It has a long history that is well worth examining more closely.

DFG has long required that all projects funded transfer their data to public data archives, for example, to the ZA or ZPID. But relatively few data sets have actually been transferred. As a result, some of the DFG’s national programmes (*Schwerpunktprogramme*, *SPP*) have imposed strict time limits for the transfer of data to public archives for every project funded. While the “success rate”—the number of projects complying with this provision—has increased, it still is far from 100%.

We may lament the discrepancy between official policy and the actual behaviour of the research community, exert more pressure, and impose tighter controls. But we also should ask: What are the reasons for this discrepancy? Why do relatively few projects “share” their data by transferring them to a data archive?

Project-specific data, generated to answer specific research questions, do not necessarily lend themselves to use by others. Both contextualisation and specification are a necessary provision for sharing these kinds of data. After completion of the research project, scarce resources—researchers’ time in particular—must be invested further to produce a data set that is potentially valuable to others and that can be transferred to an archive for their use. The question is: Can the research community’s reluctance to invest in data sharing be understood as an indicator of the low value ascribed to shared data?

And what about the data sets that *have* been transferred to archives—data from projects whose primary aim was not to produce data “for others” but to pursue specific research questions? To what degree are these data being used by the research community? In other words: Is there sufficient demand?

1 ZUMA: “Center for Survey Design and Methodology” and “Social Monitoring and Social Change”. See: <http://www.gesis.org/en/institute/gesis-scientific-sections/center-for-survey-design-and-methodology/> and <http://www.gesis.org/en/institute/gesis-scientific-sections/social-monitoring-social-change/>.

2 SOEP: German Socio-Economic Panel Study. See: www.diw.de/gsoep.

3 ZA: Data Archive and Data Analysis. See: <http://www.gesis.org/en/institute/gesis-scientific-sections/data-archive-data-analysis/>.

4 IZ: “Specialized Information for the Social Sciences” and “Information Processes in the Social Sciences”. See: <http://www.gesis.org/en/institute/gesis-scientific-sections/specialized-information-for-the-social-sciences/> and <http://www.gesis.org/en/institute/gesis-scientific-sections/information-processes-in-the-social-sciences/>.

5 ZPID is the psychology information center for the German-speaking countries. See: <http://www.zpid.de/index.php?lang=EN>.

Both of these questions—why the research community is reluctant to invest in sharing data, and how high the actual demand for shared data is—need to be analysed in more detail. Data generated with public money should, of course, be made available to the public (that is, in the case of sensitive individual or company data subject to data protection restrictions, made available to the research community). However, keeping in mind the overall goal of a data infrastructure, for some projects it may not be a top priority to invest in data sharing, given the high transaction costs and limited value of the data to the scientific community. More pragmatic approaches to secure access to individual projects' data are being discussed more in the context of “research integrity” than in the context of infrastructure.

“Providing data” is a markedly different approach, and has become more and more prominent in the DFG's funding activities over the past few years. With “data provision,” we refer to a type of project or programme whose primary aim is not to answer a specific, narrowly delimited research question and to collect data for this purpose, but to collect and/or generate data for wider use and thus act as a “research infrastructure.” The focus and theoretical foundation of this form of data production is not a set of specific hypotheses, but a wider research topic or area. Data production for wider use is the main purpose of the DFG's projects and programmes, which are designed as a service to the scientific community. Increasingly, data production is taking the form of large-scale longitudinal studies.

The DFG has long been regarded as lacking adequate funding instruments for longitudinal studies. In 1995, however, the DFG began considering how to remedy this problem, and held a workshop convening experts from the field of large-scale longitudinal studies and members of the DFG's committees. The workshop resulted in a paper that specified the criteria that would need to be fulfilled in order for longitudinal studies to seek DFG funding, and that encouraged researchers to develop their ideas for such studies.⁶ While this did not produce any significant immediate effect, the situation has changed dramatically in recent years. Large-scale longitudinal studies providing research infrastructure for the social sciences have become a major activity. Various factors have contributed to this change:

- a. Emerging activities in the national research community, closely linked to similar activities in Europe and elsewhere;
- b. Increased attention to these developments in European programmes and European institutions;
- c. Adjustments of DFG instruments to foster and promote these activities.

2. “Providing Data”—Shaping the Instruments

What did the DFG do and why? It all goes back to the workshop of 1995, where the first strategic debate took place on how the DFG could improve opportunities for funding longitudinal studies. The workshop brought together representatives from all disciplines of the social and behavioural sciences. Its recommendations addressed the scientific community as well as the DFG as a funding organisation.

This initial input did not produce systematic changes, however, either in the scientific community or at the DFG. This changed, however, with a major strategic initiative launched by the DFG in 2002, called the “Förderinitiative Geisteswissenschaften” (*Funding Initiative for the Humanities*). This initiative addressed the specific needs of the humanities, but also created new funding opportunities open to both the humanities and the social sciences. One of the four pillars of the strategic initiative was to reshape and modernize the DFG's strategic initiative “*Langfristprogramm*” (long-term programme), whose effects became visible as early as 2003.

The *Langfristprogramm* had been in existence since the DFG was founded, but was initially

⁶ The paper was widely published: *Kölner Zeitschrift für Soziologie, Psychologische Rundschau, Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, Zeitschrift für Wirtschafts- und Sozialpsychologie, and ZUMA-Nachrichten.*

designed only for the humanities. In 2003, the DFG's *Senat* and Joint Committee resolved to implement a reform of this programme with the following elements:

- a. Limits were placed on the formerly open-ended time frame: The programme is now only for research activities requiring seven to twelve years of funding.
- b. Only projects of potentially high scientific impact and importance will be funded. A longer-term perspective is necessary, but is by no means the sole requirement.
- c. The *Langfristprogramme* is no longer confined to the humanities, but is now open to both the humanities and the social sciences. The strategic decision to open up the *Langfristprogramm* to the social and behavioural sciences was based, among other things, on the recommendations from 1995. Longitudinal studies are invited to seek funding within the *Langfristprogramm*.
- d. As a consequence of provisions b and c (aiming at high-impact activities and opening up to the social sciences in general and longitudinal studies in particular), the scale of funding per individual project has been expanded: substantial funding is available depending on the individual project needs. As a consequence, fewer projects will be funded, but they will come from a broader range of disciplines—humanities and social sciences—and with a broader range of budgets.

The first project in which this new funding option was put to use was the European Social Survey, an internationally comparative study of repeated cross-sections, with more than 20 countries participating. The European Commission provides the core funding, and more than 20 national funding agencies finance the national data collection. The *Langfristprogramm* was essential in allowing the German part of the European Social Survey and the DFG to fully participate in the European programme. The DFG's decision to approve a project as part of the *Langfristprogramm* entails its commitment to provide funding for the entire duration of the activity. Because the ESS was part of the *Langfristprogramm*, the DFG was able to stand in for ESS in the network of national support institutions, the European Commission, and the European Science Foundation, and formally sign commitments. This provided the basis for the ESS as a truly European infrastructure that eventually became part of the road map of the European Strategy Forum of Research Infrastructure (ESFRI). As a consequence, ESS may become a "European Research Infrastructure," which will require a new legal form. The aim is to become a kind of international organisation. This will certainly have implications for the role of national funding organisations that are still unknown to us.

Just recently, in December 2008, the German Longitudinal Election Study (GLES) was adopted as part of the *Langfristprogramm*, with the potential to be funded for nine years. After that time and after having gathered data on three successive national elections, it is planned that GLES will be taken on board at GESIS. Whereas the future perspective for ESS beyond its funding as part of the *Langfristprogramm* remains open, the future of GLES is relatively secure: provided that the DFG-funded project proves to be a success in scientific terms, it will be continued under the institutional umbrella of GESIS.

The situation of PAIRFAM, the panel study of intimate relations and family members, is unique in another respect as well: a national research programme (*Schwerpunktprogramm–SPP*) was set up by DFG to develop and implement the study. Normally, national programmes aim at rather loose cooperation between projects around a common topic. With PAIRFAM, however, the very idea of the programme was to develop a common product. This required a clearly defined division of labour between the individual projects within the programme, a high level of coordination, clear leadership and collaboration across the field of sociological, economic, and psychological research on family and relationships. Although the funding instrument that was used—SPP—normally aims at supporting a different kind of scientific cooperation, the adaptive use of this instrument was successful, and indeed innovative: The first four years of the *Schwerpunktprogramm* were used for the development of the panel study, and the final two years are presently being devoted to carrying out the first two waves of the panel.

Before giving a “green light” for the final two years and releasing the actual funding for the first two waves, the *Senat* and Joint Committee of DFG have carefully considered the future prospects for PAIRFAM; after all, it would not have made sense to finance the first two waves without a perspective for further steps. The debate was based on a review panel’s assessment of PAIRFAM’s plans. Reviewers, *Senat* and Joint Committee came to the conclusion that PAIRFAM should be invited to seek future funding as part of the *Langfristprogramm*. This opened up a perspective of twelve years for PAIRFAM and confirmed the strategic decision to design the *Langfristprogramm* in a way that would allow substantial funding of individual projects. PAIRFAM, which started as part of a *Schwerpunktprogramm*, demonstrates that the new *Langfristprogramm* is not the only instrument in the DFG’s portfolio that can be used to support large-scale longitudinal studies.

In principle, all funding instruments should be considered. The Social-Economic Panel (SOEP) is a prominent example. SOEP, which has become a cornerstone of the German research infrastructure in the social and behavioural sciences, was initiated and developed many years ago as part of a collaborative research centre (SFB). When this SFB ended (in 1991) after an initial twelve-year funding period, SOEP’s funding was continued through the individual-project funding mode (re-financed by special funds from the German federal and state governments). However, given SOEP’s importance as a research infrastructure, an institutional solution was needed that could provide long-term stability. Between the Federal Ministry of Education and Research (*BMBF*), the respective ministries of the states and the DFG, a solution was negotiated: after more than 12 years in the individual-project funding mode) SOEP was established as a special “service unit” at DIW Berlin, which is a member of the Leibniz Association (WGL). SOEP’s success story—not in scientific terms but with regard to funding and institutional solutions—is rooted in the adaptive use of several funding instruments and cooperation among the funding institutions (BMBF, DFG, WGL). PAIRFAM, on the other hand, is currently in the process of adapting several funding instruments to its needs.

A final example of both adaptation of funding instruments and cooperation among funding institutions is the National Education Panel Study (NEPS). The idea for NEPS was first presented and discussed at the symposium in 2004 that was organised by the DFG as part of its “Programme on Empirical Research on Education”. The symposium brought together researchers from Germany and other European countries, as well as representatives of government ministries. At its conclusion, the Programme’s Scientific Board gave advice that formed the basis for the DFG’s position on NEPS. Following the Scientific Board’s recommendation, the DFG’s Governing Board agreed that the DFG would play an active role in the future process, in close collaboration with BMBF, whereby the funding for NEPS would come solely from the BMBF.

The DFG organized preparatory expert meetings, an international expert workshop to assess the pilot study, and a full-scale international peer review for the full proposal. Based on this peer-review, the BMBF made the formal decision to finance NEPS as a data-providing research infrastructure. The DFG’s *Senat* simultaneously decided to allocate a substantial budget for a national research programme (SPP) in which projects would be funded that make scientific use of the NEPS data. In other words: the DFG, by implementing its mechanisms for independent assessment of scientific quality, provided the mechanism to firmly root NEPS in the scientific community. For the implementation of the peer-review results, the BMBF and DFG agreed on a division of labour: BMBF finances the research infrastructure, DFG funds the scientific use of the data through its national research programme.

3. DFG’s Role and Contribution

Major large-scale longitudinal studies that now serve as a data research infrastructure for the social sciences have developed into a major line field of activity at the DFG. However, this was not the result of a strategic master plan. Of course, there was the policy statement of 1995 and the strategy decision of 2003 to redesign the *Langfristprogramm* specifically geared towards longitudinal studies in the social sciences.

But the individual activities and projects that emerged within the scientific community were pursued in a relatively uncoordinated way. This is not surprising, given that the DFG is owned by the scientific community and firmly founded on the principle that strategic initiatives as well as individual funding decisions must be driven by research questions and by researchers themselves. NEPS does not follow this principle to the letter, but nevertheless demonstrates the DFG's role: NEPS was initiated and in its early stage conceptualised by the BMBF. And it is the BMBF, not the DFG that funds this research infrastructure. Close cooperation and partnership with the DFG was sought to provide scientific quality control through independent peer review and thereby scientific legitimation. The DFG's role in the partnership with BMBF has been to ensure that this externally initiated panel study is and will continue to be essentially science-driven.

Closely related to the principle "science-driven" is the fact that the DFG cannot provide institutional funding, but is confined to project funding. The major strategy decision to redesign the *Langfristprogramm*, therefore, meant redefining it as a project funding mode and introducing the twelve-year limit for each cycle of funding. This means that longitudinal studies can be funded by the DFG under one of the following provisions: a) The study will come to an end within twelve years. b) The topic of the study demands a longer perspective than twelve years, but if no continuation can be secured, the scientific outcome of twelve years alone will justify the investment. In other words, the second-best solution can be a stand-alone one. c) The study is planned from the outset as a truly longitudinal one, going beyond twelve years: Initial funding by the DFG can be granted if the follow-up, i.e., institutional funding can be expected. SOEP (which was not planned as such a long-term project, but in fact became one) and GLES (which was planned as such from the very beginning) are examples of the DFG strategy of enabling a potentially long-lasting project to be launched. This brings us to our first conclusion regarding the "role and contribution" of the DFG:

Projects that seek funding from the DFG have to be driven by the scientific community, i.e., they must be well planned scientifically, and they must be organised in a form that is suitable to the project funding—at least for the duration of the DFG funding. If these two provisions are fulfilled, the DFG is well equipped to find adaptive solutions.

Projects like ESS, GLES, and PAIRFAM are data research infrastructures of central importance to the research community; yet they are also expensive and put considerable strain on budgets available for funding the social sciences. Up to now, these projects have been proposed individually and dealt with on a case-by-case basis. However, if data research infrastructures are going to establish themselves as a major line of activity and funding, some degree of coordination and even strategy might be necessary. The DFG's elected bodies—*Fachkollegien* and *Senat*—will be able to provide leadership for this process of addressing key questions within the scientific community: mapping the field, defining thematic priorities, coordinating projects and programmes in order to maximise effects and economise resources, etc..

Coordination and strategy also pose challenge to the DFG as a funding organisation, to the ministries, and to research organisations like the Leibniz Association (WGL) and universities. Coordination and collaboration between the institutions have up to now also taken place on a case-by-case basis: SOEP (DFG/BMBF/WGL), GLES (DFG/WGL) and NEPS (DFG/BMBF/University) have each resulted in individual constellations and solutions that we regard as success stories. But again:

We have witnessed increased activities in this field, and the momentum has been building. Not only because of the financial implications, but also in view of the long-term perspective of each individual activity, coordination and collaboration between the major players in the field may become necessary. A division of labour and development of institutional perspectives are the keywords here. The DFG is prepared to play an active role in this coordination process.

Coordination and collaboration between institutions is not only appropriate in view of the division of labour and sharing of responsibilities on the national level, but also in view of the international activities. ESFRI is but one field, however important it may be. If “European Research Infrastructures” come into existence as new legal entities, we as national institutions will have to define our position vis-à-vis these new entities as well as in relation to each other. The national institutions will have to cooperate in order to maximise the effects on the European and international level—and of course, in the best interests of the research community.

The International Data Forum (IDF), finally, goes beyond the European level. The DFG has supported the initial phase of the idea, together with our partner organisations from the UK, the US, Canada, the Netherlands and China. The IDF aims to facilitate and coordinate international production and sharing of data for research in the social sciences. It strives to align its aims with the strategic directions and priorities of prominent organisations representing the producers, managers and research users of data relevant to the social sciences. One of its tasks is to facilitate collaboration and mutual understanding between key data stakeholders in the social sciences. Since the founding conference for the International Data Forum, the next steps are to establish interagency agreement on the need for IDF and the scale of its operations. Decisions will be sought in 2009.

DFG has nominated the chair of the *RatSWD* (German Council for Social and Economic Data) as a member and the German representative of the founding committee of the IDF. This, already, is a concrete example of coordination between national institutions.

4. Summary

The DFG’s strategy for data research infrastructures will be based on what has already been achieved and on the lessons that can be learned:

1. “Providing data,” rather than “sharing data,” should be the guiding perspective.
Projects whose primary purpose is to provide a common good should focus on building research infrastructure.
2. The DFG has powerful programmes at its disposal to fund outstanding infrastructure projects. It is up to the scientific community to adapt the DFG’s diverse funding instruments to its needs.
3. Strategic cooperation is needed among all interested parties: cooperation within the research community on thematic priorities; cooperation between the research community and funding institutions on options for funding; and cooperation between the funding institutions on the division of labour, on the national as well as on the international level.
4. The DFG is prepared to play an active role in this cooperation under the leadership of its elected bodies (*Fachkollegien* and *Senat*).