German Council for Social
and Economic Data (RatSWD)

www.ratswd.de

RatSWD
Working Paper Series

Working Paper No. 68

Interdisciplinary Longitudinal Surveys: Linking Individual Data to Organizational Data in Life-course Analysis

Stefan Liebig

March 2009

Federal Ministry
of Education
and Research

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 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.

Interdisciplinary Longitudinal Surveys: Linking Individual Data to Organizational Data in Life-Course Analysis

Stefan Liebig

I. Research Questions

One of the fundamental insights gained by the social and economic sciences is that empirically founded statements on the conditions and consequences of individual behavior or of social and economic change can only be formulated on the basis of longitudinal microdata. The observation of individuals, households, and other socio-economic units over long periods of time allows us to causally determine the reasons for social and economic stability and change. Moreover, socio-economic phenomena are particularly path-dependent. The opportunities and restrictions that individual or corporate actors face over their life courses — or more generally: over time — depend to a great extent on decisions and events earlier in time. The available individual and household-level datasets used in empirical social and economic research in Germany are capable of mirroring these path-dependencies.

But social and economic phenomena show another fundamental quality: they are embedded in social contexts and social aggregates (Granovetter 1985). Embeddedness means that actors are in most of the cases elements of a number of social aggregates. Their behavior is affected by these different memberships and the structures and processes that take place within these aggregates, whether households, social networks, schools, firms, associations, regional areas, or nations. Longitudinal microdata for assessing the effects of these different social contexts on individual decisions and behavior are available at the level of households, geographic units, or — within comparative research — at the national level. However, the recent labor market and educational research shows that there is another type of social aggregate that is crucial for an individual's economic or social situation and his or her life chances: institutions and organizations like schools, universities, firms, or establishments (Baron/Bielby 1980, Coleman 1993, Hamermesh 2008, Heckman 2001).

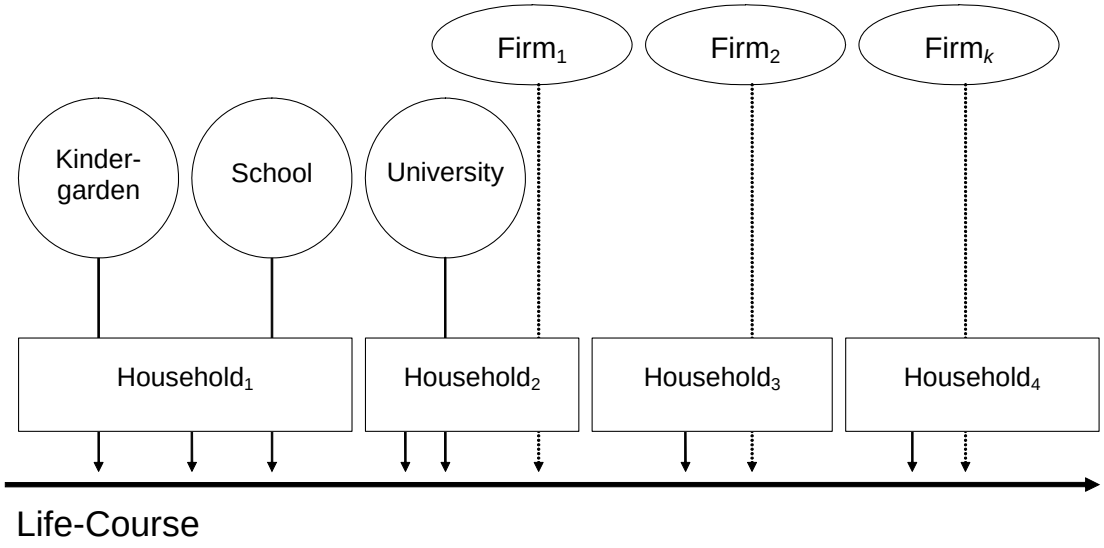
For many years, the level of organizations has played a subordinate role in the German research on social stratification, the labor market, and the education system (Allmendinger/Hinz 2002). With regard to firms and establishments, this was justified with

reference to the dominance of the tariff system and the longstanding practice of macro-level regulation. Today, there exist a range of empirical studies showing a general trend towards increasing heterogeneity on the organizational level in Germany and suggesting that labor market and educational institutions are developing more and more differentiated internal structures and processes. One consequence of this development has been that the distribution of goods, jobs, and life opportunities is determined increasingly by the "internal logic of organizations." Some of the main effects of this on individual career paths and employment histories can be seen at the establishment and firm level (Bender et al. 2000, DiPrete et al 2001), on the level of wages (Kölling et al. 2005), in the duration of unemployment and in qualification levels (Frederiksen et al. 2006), and even in the political attitudes of employees (Liebig/Krause 2007). Besides the classical variables such as number of employees (Heyman 2007), degree of unionization (Fitzenberger et al. 2007), and branch affiliation, a range of other important explanatory factors can be identified on the firm and establishment level, such as a firm's age (Brixy et al. 2007), its socio-demographic structure (Krell/Sieben 2007), the magnitude of income disparities or mobility chances (Liebig/Krause 2007), and the particular form of work organisation (Bellmann/Pahnke 2006).

The operative structures, processes, and strategies, as well as the business situations of employers are becoming increasingly important, and not only for employment revenues (Goedicke 2006, Lengfeld 2007). The variety of firm-specific operative time regimes, improvements in the compatibility between work and family, health promotion activities, and more flexibile regulations governing working time and location (e.g., home workplaces) also affect an individual's social relations and his or her way of life in general (Düntgen/Diewald 2007).

As has already been outlined, organizations can control their members' access to jobs and goods. This is an assumption that takes on particular importance when analyzing durable structures of social inequality (Tilly 1998). The individual life course can also be understood as a sequence of different memberships in organizations (Figure 1). Individual life courses can thus be distinguished by the extent to which people succeed in joining organizations that offer better life chances. In this context, social stratification research tries to investigate whether this also results in path-dependencies, i.e., as people become members of advantageous or disadvantageous organizations, advantages and disadvantages are accumulated over the life course.

Figure 1: The individual life course and memberships in different types of organizations



2. Status quo

In order to empirically analyze the effects of the organizational level on individual career paths, on the conditions and outcomes of employment, as well as on different aspects of individual life courses, social and economic research requires adequate data linking personal and organizational information. Such matched organization-member datasets are available especially in the field of labor market research. These linked employer-employee (LEE) data sets are characterized by a hierarchical multi-level structure, in which employees constitute the bottom level and the firms and/or enterprises constitute the upper level. The distinct feature of these LEE data is that they contain information about several — and in the optimal case, all — persons employed in a firm. In most cases, "process-produced" administrative data, on either the individual or the firm level, constitute the basis of analysis (Abowd/Kramarz 1999). In contrast to other European and non-European countries, Germany recognized the potential of LEE-data very late. This is why, in 2001, Martin Falk and Viktor Steiner concluded, in their expert report to the Commission for the Improvement of the Informational Infrastructure between Science and Statistics (KVI): "The opportunities of matching firm and individual data were recognized much earlier in other countries. In certain areas, such as operative employment and income trends, German research is no longer competitive. In this domain, research is almost non-existent" (p. 8).

In the meantime, the data supply has been improved substantially, mainly because of the linked employer-employee dataset of the IAB (LIAB) (Alda et al. 2005) and the income and wage structure surveys conducted by the official statistical agencies (Stephan 2001), which are available in the research data centres of the Federal Employment Agency and the Statistical Offices of the Federal Statistical Office and the Statistical Offices of the Federal States. Both data sources are "real" linked employer-employee datasets that offer information on all — or at least a sufficient number of — employees in each participating firm. Both datasets contain vast and diverse potential for analysis. The central difference is the degree of available firm information contained. The income and wage structure survey is a cross-sectional dataset; it only contains the basic parameters of the employment structure, sectoral affiliation, and degree of collective bargaining. Thus it can be used primarily for the analysis of cross-sectional wage structures (especially after the inclusion of surveyed firms and sectors through changes in the legislation in January 2007). The LIAB, on the other hand, offers a broader base of information, ranging from detailed employment structures, the firm's economic situation, professional training programs, to labor time regulations, payment systems, and special measures to improve compatibility between work and family. Although this focus indeed requires further development — e.g., with regard to the existing mobility regimes or the firm culture, which is quite important to organisational research — on the operative side, the LIAB offers a potential for analysis that exceeds the classic labor economic or sociological questions, all the more so because it displays longitudinal processes on the firm and on the individual level. This central advantage is diminished, however, by the restricted supply of information on the employee side. Here, the LIAB shares one of the main weaknesses of the income and wage structure survey.

Both available LEE data sets in Germany are characterized by restricted access to information on individuals and households. This applies to central features of current employment relationships (the LIAB does not identify, e.g., temporary employment or the supply of temporary workers), to information on the economic situation of an individual, and even more so to household data, the family situation, social origins, social preferences and personal characteristics, norm and value orientations, and political attitudes and membership in parties or other organizations. Since these topics are of central interest in empirical social and economic research, there is a strong need for a dataset that contains longitudinal information on the individual, household, and organizational level.

Against this background, an extension of the existing linked employer-employee data supply in Germany is desperately needed. This improvement needs to be promoted especially

for the kind of information that goes beyond basic employment data. This can be achieved, for instance, by gathering information on family background, family and domestic situations, integration into social networks, as well as moral concepts and political attitudes. Improvements are also possible on the organizational side — the data catalog of the IAB establishment panel can, for instance, be expanded to include income and wage formation processes, elements of enterprise and firm culture, industrial relations, and the national or international competitive position of firms. Such a catalog of information can only be created on the basis of linked employer-employee surveys. In the present research, these kinds of data are produced using two different approaches:

(1) Employer-First Approach

In the first step of this approach, which has also been pursued by official statistics in the framework of the income and wage structure survey or the WeLL project by IAB and RWI (Bender et al. 2008), suitable firms are selected. Individual information is collected from a sample of employees working in these firms (either all employees or a partial sample) (cf. the 2000 National Employer Survey, Capelli 2001). The advantage here is that the existing multi-level data structure prevailing in common LEE data sets is still existent. One problem, however, is that such samples quite rapidly go beyond realistic limits. This happens if the information on the employees is not supplied by the firm itself but gathered by employee surveys. The coordination and implementation of such employee surveys in more than 100 or 200 firms is hardly practicable in the framework of normal research projects — even when the surveys are conducted by survey institutes. Accordingly, a recent project in Germany utilizing this approach concentrated on a single-digit number of firms (Goedicke et al 2007).

(2) Employee-First Approach

In the second approach to generating matched datasets, not firms or organizations but persons (employees) constitute the point of departure. The individual data, which are gathered through personal interviews, are later complemented by firm data. This again can be done in three different ways (a technique that is already being used in research projects) (cf. Kmec 2003):

(1) The information on the establishment or firm where the respondent of a population survey, e.g., is employed are added using available commercial business datasets (in Germany: *Creditreform* or *Hoppenstedt*). Examples of this method are the New Worker Establishment Characteristics Database and the Decennial Employer-Employee Dataset. The problem of this approach is the limited scope of available firm information in the databases

(e.g., number of employees, founding year, business volume). Although business databases can be used to assess an enterprise's liquidity or financial strength, they are less suitable for scientific questions.

(2) The second way is complement the personal information by establishment or firm data from official statistics. In the framework of a study conducted by the Max Planck Institute for Human Development in Berlin, for instance, researchers asked the respondents for their social security numbers. Afterwards, the individual data were linked to the IAB Establishment Panel (Reimer/Kuenster 2004). If the employer was included in the IAB establishment panel, the firm information was added to the individual data record. Obviously the problem herein is that the share of employees in a population survey who are covered at the same time by the IAB establishment panel is expectedly small. Another possibility lies in using the IAB Establishment History Panel (Dundler et al. 2006), but in this case, the available employer information is much more restricted than in the IAB Establishment Panel.

(3) Finally one can use an individual or household survey to ask employees for the name and address of their employer, and can conduct a separate firm survey on the grounds of this information. The collected firm data can than be matched with the individual or household data. Examples of this approach are the Multi-City Study of Urban Inequality and most notably the National Organization Survey (NOS) from the years 1991 and 2002 (Kmec 2003). In the framework of the General Social Survey (GSS) of 1991 and 2002, all (1991) and, respectively, some (2002) of the currently employed were asked for the name and address of their workplace. Local business units in which people were gainfully employed were the target units. On the basis of these entries, telephone interviews were conducted and postal questionnaires distributed. These data were matched to the individual data of the GSS. The result is a linked employee-employer data set (Kalleberg et al. 1996, Smith et al. 2004). In 1991, for a total of 51 percent of all cases (in 2002, 48 percent) the individual and firm data could successfully be matched. In contrast to the classical LEE data structure, this dataset does not possess a hierarchical structure. For one firm, the individual data are available for just one employee. Due to its cross-sectional character, this does not offer causal or longitudinal potentials for analysis. But through combined individual-firm surveys, it is possible to collect far more firm information than in a person-to-person interview, and the firm data, which are collected in combined surveys, are gathered independently from the interviewee's attitudes and perceptions (cf. Gupta et al. 2000).

In a current project underway at the University of Bielefeld, the design of the NOS study is

being tested for its transferability to the German situation. For this purpose, all currently employed persons who are being surveyed in the ALLBUS 2008 (a nationwide reference survey) were asked for the name and addresses of their employers. Useful data is available for about 85 percent of those people who are employed in firms with more than six employees. On the basis of these data, a firm survey will be conducted in January 2009. The aim of this study is to assess the quality and methodological problems arising in connection to the generation of survey-based LEE-data sets. Moreover, conclusions for future interview projects will be derived. As the willingness to participate in firm surveys has decreased constantly since the 1990s, another important task will be to find ways to maximize firm participation. A central problem of such a twofold survey-based approach is data protection. The respondents have to give permission for their firms to be contacted. Only then can individual and firm information be matched. A further problem is the re-identification of individuals and firms. However, the projects currently carried out by the official statistical agencies on the anonymization of firm and panel data already offer suitable tools that simplify data access — also for researchers.

3. Future Developments

Empirical research has shown that the significance of different forms of workplace organization, labor market processes, social stratification, and other socio-economic phenomena (e.g., work-life balance) is increasing in Germany. From this follows an increasing demand for socio-economic data sets that identify linkages between individuals and organizations. Especially in the field of educational research, the interest in particular educational institutions will increase in the near future (Klieme 2008). The efficiency and the evaluation of activities will be measured according to their impact on the student's performance and his/her educational achievements. However, if no further household information is available, the linkages between organizational and individual data are not sufficient — especially with regard to the educational system.

The linkages between different data sources that can be facilitated by official statistics (Bender et al. 2007) offer the chance to broaden the scope of survey-based organizational data and to match them with information from other data sources. This reduces interview costs and allows the researcher to conduct firm surveys that are more strongly focused on a specific topic. As the socio-economic research has recognized the need for longitudinal data and the embededdness of individual behavior, it seems to be more important than ever before to

collect longitudinal information on the individual and the household level.

4. Recommendations

Against this background the following recommendations can be made:

1. There is an increasing demand for linked data between individual, household and organization information—especially with regard to the organization of the educational system and the workplace.
2. As the available data sets only offer limited information, household and individual surveys should be matched based on adequate organizational data. This can be achieved by matching data from official statistics or from separate surveys.
3. Linked individual/household and organizational data sets will be only feasible for socio-economic research if they contain longitudinal information.
4. The best solution to achieve an adequate data structure is to enrich the Socio-Economic Panel (SOEP) with separate firm surveys (e.g., of nursery schools, schools, workplaces of other household members) at five-year intervals. Respondents to SOEP should be asked for the names and addresses of these organizations, and based on this information, organizational surveys should be conducted to achieve a three-level hierarchical and longitudinal data set. In this way longitudinal information would be made available on the individual, the household, and the organizational level. Such a data set would be internationally unique and would offer a novel potential for analysis in a variety of disciplines (education, sociology, economics, psychology).

References:

Abowd, John M. and Francis Kramarz, 1999: The analysis of labor markets using matched employer-employee data. Pp. 2629-2710 in Orley Ashenfelter & David Card (ed.), Handbook of Labor Economics. Elsevier.

Abowd, John M., John Haltiwanger and Julia Lane, 2004: Integrated Longitudinal Employer-Employee Data for the United States. American Economic Review 94: 224-229.

Alda, Holger, Stefan Bender and Hermann Gartner, 2005: The linked employer-employee dataset of the IAB (LIAB). IAB Discussion Paper No. 6/2005.

Allmendinger, Jutta and Thomas Hinz, 2002: Perspektiven der Organizationssoziologie. Pp.9-28 in Jutta Allmendinger and Thomas Hinz (eds.), Organizationssoziologie. Sonderheft 42/2002 der Kölner Zeitschrift für Soziologie und Sozialpsychologie. Opladen: Westdeutscher Verlag.

Baron, James N. and William T. Bielby, 1980: Bringing the Firms Back, Stratification, Segmentation, and the Organization of Work. American Sociological Review 45: 737-765.

Bellmann, Lutz and Andre Pahnke, 2006: Auswirkungen organisatorischen Wandels auf die betriebliche Arbeitsnachfrage. Zeitschrift für ArbeitsmarktForschung 39: 201-233.

Bender, Stefan, Dirk Konietzka and Peter Sopp, 2000: Diskontinuität im Erwerbsverlauf und betrieblicher Kontext. Kölner Zeitschrift für Soziologie und Sozialpsychologie 52: 475-499.

Bender, Stefan, Joachim Wagner and Markus Zwick, 2007: Kombinierte Firmendaten für Deutschland. FDZ Methodenreport 5/2007. Nürnberg: Bundesagentur für Arbeit.

Bender, Stefan, Michael Fertig, Katja Görlitz, Martina Huber, Stefan Hummelsheim, Petra Knerr, Alexandra Schmucker und Helmut Schröder, 2008: WeLL - Berufliche Weiterbildung als Bestandteil Lebenslangen Lernens: Projektbericht. FDZMethodenbericht 05/2008. Nürnberg: Bundesagentur für Arbeit.

Brixy, Udo, Susanne Kohaut and Claus Schnabel, 2007: Do Newly Founded Firms Pay Lower Wages? First Evidence from Germany. Small Business Economics 29: 161-171.

Capelli, Peter, 2001: The National Employer Survey: Employer Data on Employment Pratices. Industrial Relations 40: 635-647.

Coleman, James S., 1993: The Rational Reconstruction of Society. American Sociological Review 58: 1-15.

Diewald, Martin, Hanns-Georg Brose and Anne Goedicke, 2005: Flexicurity im Lebenslauf. Wechselwirkungen zwischen pluralen Lebensformen und betrieblichen Beschäftigungspolitiken. Pp.223-248 in Martin Kronauer and Gudrun Linne (eds.) Flexicurity. Die suche nach Sicherheit in der Flexibilität. Berlin: Edition Sigma.

DiPrete, Thomas A., Dominique Goux, Eric Maurin and Michael Tahlin, 2001: Institutional Determinants of Employment Chances: The Structure of Unemployment in France and Sweden. European Sociological Review 17: 233-254.

Dundler, Agnes, Melanie Stamm and Silke Adler, 2006: Das Betriebs-Historik-Panel. FDZDatenreport 3/2006.

Düntgen, Alexandra and Martin Diewald, 2007: Auswirkungen der Flexibilisierung von Beschäftigung auf eine erste Elternschaft. Pp.213-231 in Mark Szydlik (Ed.): Flexibilisierung. Folgen für Arbeit und Familie Wiesbaden: VS Verlag.

Falk, Martin and Viktor Steiner, 2001: Expertise für die Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik im Bereich Erwerbstätigkeit und Einkommen, in: Kommission zur Verbesserung der informationellen Infrastruktur zwischen Wissenschaft und Statistik (eds.), Wege zu einer besseren informationellen Infrastruktur. Baden-Baden: Nomos.

Fitzenberger, Bernd, Karsten Kohn and Alexander C. Lembcke, 2007: Union Density, Collective Bargaining, and Individual Coverage: The Anatomy of Union Wage Effects. Mannheim: ZEW.

Frederiksen, Anders, Rikke Ibsen, Michael Rosholm and Niels Westergaard-Nielsen, 2006: Labor Market Signalling and Unemployment Duration: An Empirical Analysis Using Employer-Employee Data. IZA Discussion Paper No. 2132.

Goedicke, Anne, 2006: Organizationsmodelle in der Sozialstrukturanalyse: Der Einfluss von Betrieben auf Erwerbsverläufe. Berliner Journal für Soziologie 16: 503-524.

Granovetter, Mark, 1985: Economic Action and Social Structure: The Problem of Embeddedness. American Journal of Sociology 91: 481-510.

Gupta, Nina, Jason D. Shaw and John E. Delery, 2000: Correlates of response outcomes among organizational key informants. Organizational Research Methods 3: 323-347.

Hamermesh, Daniel, 2008: Fun with matched firm-employee data: Progress and road maps. Labor Economics 15: 663-673.

Heckman, James J., 2001: Micro Data, heterogeneity, and the evaluation of public policy: Nobel lecture. Journal of Political Economy 109: 673-748.

Heyman, Fredrik, 2007: Firm Size or Firm Age? The Effect on Wages Using Matched Employer-Employee Data. Labor 21: 237-263.

Kalleberg, Arne L., David Knoke, Peter V. Marsden, Joe L. Spaeth (Hrsg.), 1996: Organizations in America. Thousand Oaks: Sage Publications.

Klieme, Eckhard, 2008: Längsschnittuntersuchungen zur Schulentwicklung. Vortrag auf der 4. Konferenz für Sozial- und Wirtschaftsdaten. (http://www.ratswd.de/kswd/download/4_KSWD_Klieme.pdf, 20.09.2008).

Kmec, Julie A., 2003: Collecting and Using Employer-Worker Matched Data. Sociological Focus 36: 81-96.

Kölling, Arnd, Claus Schnabel and Joachim Wagner, 2005: Establishment Age and Wages: Evidence from German Linked Employer-Employee Data. Pp. 81-99 in Lutz Bellmann, Olaf Hübler, Wolfgang Meyer and Gesine Stephan (eds.), Institutionen, Löhne und Beschäftigung, Beiträge zur Arbeitsmarkt- und Berufsforschung Nr. 294. Nürnberg: IAB.

Krell, Gertraude and Barbara Sieben, 2007: Diversity Management und Personalforschung. Pp.235-254 in Gertraude Krell, Barbara Riedmüller and Barbara Sieben (eds.), Diversity Studies - Grundlagen und disziplinäre Ansätze. Frakfurt am Main: Campus.

Lengfeld, Holger, 2007: Organisierte Ungleichheit. Wie Organisationen Lebenschancen beeinflussen. Wiesbaden: VS Verlag für Sozialwissenschaften.

Liebig, Stefan and Alexandra Krause, 2007: Arbeitsorganisationen als Kontexte der Einstellungsbildung. Pp.15-48 in Hans J. Hummell (eds.), Die Analyse von Gesellschaften, Organisationen und Individuen in ihrem Zusammenhang – Theoretische und methodische Herausforderungen. Tagungsberichte Band 13, Bonn: Gesis.

Reimer, M. and R. Kuenster, 2004: Linking Job Episodes from Retrospective Surveys and Social Security Data: Specific Challenges, Feasibility and Quality of Outcome. Arbeitspapier Nr. 8 des Projekts Ausbildungs- und Berufsverläufe der Geburtskohorten 1964 und 1971 in Westdeutschland. Berlin: Max-Planck-Institut für Bildungsforschung.
Schnabel, Claus, 2006: Verbetrieblichung der Lohnfindung und der Festlegung von Arbeitsbedingungen. Arbeitspapier 118. Düsseldorf: Hans-Böckler Stiftung.
Smith, Tom W., Arne L. Kalleberg, Peter V. Marsden, 2004: National Organization Survey (NOS), 2002. Ann Arbor: ICPSR.
Tilly, Charles, 1998: Durable Inequality. Berkeley: University of California Press.