

Problems of linking theory and data in historical sociology and longitudinal research

Baur, Nina

Veröffentlichungsversion / Published Version

Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:

GESIS - Leibniz-Institut für Sozialwissenschaften

Empfohlene Zitierung / Suggested Citation:

Baur, N. (2009). Problems of linking theory and data in historical sociology and longitudinal research. *Historical Social Research*, 34(1), 7-21. <https://doi.org/10.12759/hsr.34.2009.1.7-21>

Nutzungsbedingungen:

Dieser Text wird unter einer CC BY Lizenz (Namensnennung) zur Verfügung gestellt. Nähere Auskünfte zu den CC-Lizenzen finden Sie hier:

<https://creativecommons.org/licenses/by/4.0/deed.de>

Terms of use:

This document is made available under a CC BY Licence (Attribution). For more information see:

<https://creativecommons.org/licenses/by/4.0>

Problems of Linking Theory and Data in Historical Sociology and Longitudinal Research

*Nina Baur**

Abstract: »Probleme beim Verknüpfen von Theorie und Daten in der Historischen Soziologie und Längsschnittforschung«. Theory and data are closely linked in empirical research: Data are the main source for building and testing theories, and without theoretical focus, it is impossible to select and interpret data. Still, the relationship between theory and data is only rarely discussed and, if so, only on a general level. Focussing on process-oriented and longitudinal research questions, the authors of this special issue contribute to this discussion by elaborating some data types that can be used for analyzing long-term social processes. For each specific data type, it is important to ask about their specific characteristics and how this effects interpretation. The authors address these questions from a broad range of theories and by either re-analyzing research-elicited data or by using process-generated data.

Keywords: Theory, Data, Empirical Research, Longitudinal Research, Process-Generated Data, Research-Elicited Data; Historical Sociology; Historical Social Research.

1. The Importance of Discussion the Relation of Theory and Data

Theory and data are closely linked in empirical research: Data are the main source for building and testing theories, and without theoretical focus, it is impossible to select and interpret data (Knoblauch 2008). Without having a concept of how theory and data are linked, it is impossible to assess the validity of research.

Despite this close link, there is remarkable little discussion one the relationship between theory and data/sources, and if so, usually on a very general level. Novices to the fields of history and sociology even might get the impression that theory and data are distinct, unrelated concepts. In a way, one could argue that the strict division into theory and data was one of the reasons for disciplinary divide between sociology and history during the 19th century, with early sociologists stressing the importance of theory (and a tendency of neglecting

* Address all communications to: Nina Baur, Institut für Soziologie, Technische Universität Berlin, Sekretariat FR 2-5, Franklinstr. 28/29, 10587 Berlin, Germany; e-mail: nina.baur@tu-berlin.de.

data) and historians stressing the importance of data quality (and a tendency of neglecting theory).

At least in German sociology, this tendency was replicated within the discipline after World War II, so today one could argue that sociological theorists generally discuss the appropriateness of different concepts which they do not test themselves. Neither do they discuss which methods would be most appropriate for testing theories. The modern discourse on methods of social research generally covers questions like data selection, collection and analysis. However, the question as to how we can derive a research question from a theoretical problem and operationalize it is dealt with briefly at the most. The question as to how get from data to sociological theories is not covered at all (Baur 2008a).

In other words: Theorists usually talk about theories, methodologists talk about data. As data are the main source for building and testing theories, and as it is impossible to select and interpret data without theoretical concept, researchers have to tackle many of the “really important” methodological questions in research practice without having systematic guidelines for doing so. Thus a discussion on linking theory and data is overdue.

This does not mean that there has never been a discussion on the relationship between theory and data. There is, on the contrary, a broad epistemological discussion on the issue. One of the traditional ways of tackling the problem is the distinction between deduction, induction and abduction. Closely linked to this issue is the question, if research should rather test theories or build theories and how the research process has to be arranged in order to achieve this goal. Finally, there has been a fierce discussion within sociology, if overall validity is better, if the research process is linear or circular/iterative.

As important as these discussions are, they usually remain on a rather abstract level and can only provide general guidelines. Everyone who has actually done empirical research in practice knows that many puzzles concerning the relationship between theory and data remain unsolved. This special issue aims at re-opening this discussion, at pointing out some open questions and – hopefully – closing some.

At first site, the articles in this special issue are very distinct. Contributors not only stem from both history and sociology, but, as table 1 illustrates, also stem from a broad range of thematical fields ranging from family over military to media research. Authors also widely vary in the historical phase and geographical area their research addresses: Most contributors to this special issue address research questions concerning Germany, but there are also paper on Australia and the United Kingdom (Young), Austria (Kuzmics, Mayer), Canada (Freund), the U.S.A. (Jones/Peskin) and on global developments (Young). Some authors address very long-term social processes (Ernst, Franke, Jones/Peskin, Kuzmics, Mayer, Young), others specific historical periods, e.g. National Socialism/World War II (Freund, Rass), post-war developments (Bau-

ernschmidt, Hagenah/Meulemann, Salheiser, Volkens et al.) and developments since 1989 (Baumgarten/Grauel, Schmitz et al.).

In a way, I believe, this disparity is exactly what is needed: In many theoretical subfields, researchers typically do research using a specific theoretical frame and a specific type of data. For example, in biographical research in German sociology, most researchers use the theoretical framework and methodology provided by Fritz Schütz' technique of narrative interviewing. In social structural analysis and labour market research, approaches usually use standardized data and have moved to life-course-perspectives in recent decades. This also means, that there is vastly accumulated knowledge concerning the linkage of theory and data for specific theories and specific data which is generally known within the thematic subfield, but rarely known in other subfields. Drawing together contributions from different disciplines and subfields is the first step for methodological transfers between these disciplines and subfields.

Table 1: Thematic Topics Addressed in Articles

Topic	Authors
Education	Franke, Salheiser
Elites	Baumgarten/Grauel, Franke, Kuzmics, Salheiser
Emotions	Kuzmics
Family	Franke, Jones/Peskin, Schmitz et al.
Gender	Ernst, Jones/Peskin
Health	Jones/Peskin
Life Courses and Biographies	Franke, Rass, Salheiser
Media, Culture and Communication	Bauernschmidt, Baumgarten/Grauel, Hagenah/Meulemann
Migration, Ethnicity and Racism	Freund, Mayer, Young
Military	Franke, Kuzmics, Rass
National Socialism / World War II	Freund, Rass
Organisations (e.g. Military, Parties)	Mayer, Rass, Volkens et al.
Politics	Baumgarten/Grauel, Volkens et al.
Time Use	Hagenah/Meulemann
Work, Careers, Labour Market	Ernst, Franke, Salheiser

At the same time, despite their disparities in thematical issues, historical period and geographical area addressed, all contributors to this special issue have one common theoretical theme: They focus on process-oriented and longitudinal research questions and elaborate data types that can be used for analyzing

long-term social processes. Additionally, there are at least three ways of how the debate between authors in this special issue can be structured:

- 1) Which *type of theory* do authors address and how does this effect possible links to data?
- 2) Which *type of data* do authors use and how does this effect possible links to theory?
- 3) Which *phase of the research process* do authors address and how are theory and data linked within this phase of the research process?

2. Type of Theory

One of the most recent contributions to linking theory and data in German sociology is “Theoretische Empirie” (Kalthoff et al. 2008). In the introduction Kalthoff (2008) points out that the first problem in discussions on the relationship between theory and data is defining what theory actually is. Lindemann (2008) points out that theories differ in their level of abstraction. She suggests distinguishing at least types of theories:

- 1) Social Theories (“Sozialtheorien”) contain general concepts about what society is, which concepts are central to analysis (e.g. actions, interactions, communication), what the nature of reality is, what assumptions have to be made in order to grasp this reality (and this to be able to do empirical research at all) and how – on this basis – theory and data can be linked on a general level.
- 2) Middle-range theories (“Theorien begrenzter Reichweite”) concentrate on a specific thematic field (such as listed in table 1), a historical period and a geographical region. They model social process just for this sociohistorical context. For example, Esping-Andersen’s (1990) model of welfare regimes argues that there have been typical patterns of welfare development in Western European and Northern American societies since about the 1880s. In contrast, in their study “Awareness of Dying” (1975), Glaser and Strauss address topics of medical sociology and claim to have identified typical patterns that are valid for the U.S. in the 1960s and 1970s.
- 3) Theories of Societies (“Gesellschaftstheorien”) try to characterize complete societies by integrating results from various studies to a larger theoretical picture (e.g. Capitalism, Functionally Differentiated Society, Modernity, Postmodernity). In other words, they build on middle-range theories and further abstract them.

As can be seen from table 2, authors in this special issue use a broad range of social theories (as different as Rational Choice Theory, Figurational Sociology, Biographical Research, Discourse Theory and Cultural Theories). These theories are the frame for selecting specific data and for building middle-range theories. In other words: Almost all authors use this frame for arranging the research-process in order to build and test middle-range theories. They discuss

data also from these specific theoretical standpoints, and by cross-reading articles, one can see how links between theory and data differ depending on the general social theory used.

Table 2: Social Theory Used by Authors

Social Theories	Authors
Biographical and Life Course Theory	Freund, Jones/Peskin, Rass
Cultural and Social Memory	Freund
Discourse Theory and Framing	Baumgarten/Grauel
Figurational and Process Sociology (Elias)	Ernst, Kuzmics
Network Theory	Mayer
Rational Choice Theory	Schmitz et al.
Saliency Theory	Volkens et al.
Social Class and Social Inequality	Franke, Salheiser, Rass, Young
Symbolic Cultural Theory (Geertz)	Bauernschmidt
Value Theory	Hagenah/Meulemann

In the last issue of HSR (33.4), I have argued that social scientists need a common framework which helps comparing social and middle-range theories with data and specific research practice (Baur 2008a). I have suggested that such a framework should consist of at least sub-dimensions, which of course have to be filled with content: (1) Action Sphere; (2) Analysis Level; (3) Space; (4) Time with the two sub-dimensions (4a) duration and (4b) pattern. If one applies this framework to the articles in this special issue, it becomes (as already stated above) obvious that they vary in action sphere, analysis level, spatial unit address and in duration. However, they have in common that their main theoretical focus is exploring patterns of social processes in time. Thus, the contributions to this special issue can be read applying the overall question: What can specific data type contribute to historical sociology and to analyse long-term social processes?

3. Which Data Type for which Theory?

So far, I have discussed addressing the linkages of theory and data starting from the question of which theoretical problems are addressed and how this influenced the handling of data. Another way of tackling the relationship between theory and data is from the data:

Historians have for long time pointed out, that different data can be read in different ways. In other words: How to interpret data and which kind of information can be drawn from data depends on the interpreters' perspective. This

perspective in turn depends, first, on the researchers' specific historical and geographical origin – each epoch has its own pressing problems and its own important new questions. However, if a researcher is not aware of her perspective, there is a danger of interpretation being normative and biased, as the example of 19th century German historian shows: Although (or maybe exactly because) they wanted to be “neutral”, they unconsciously told history from the perspective of the powerful. Additionally, most German historians were national-conservative, anti-democrat and Eurocentric. Consequently, early German historical science served to politically legitimate historically evolved orders (Wehler 1980: 8, 44; 53-54).

Early German sociologists responded to this way of doing history by pointing out that theory can be a way of making one's perspective explicit and by reflecting it. In other words: Depending on one's theoretical standpoint, the same data give answers to different questions. Still the question remains (and this is another way of reading this special issue): Can really all data answer all types of theoretical questions?

The simple answer is: no. Or more precisely: It seems, that some data are better suited for some questions, lesser for others. For example, one can read in most German introductions to methods of social research that oral data (e.g. qualitative interviews, survey data) are better suited for grasping people's thoughts, opinions, values and inner feelings, while observation and ethnography are better suited for addressing people's actions. Of course, from seeing what people do, one can try to conclude why they did it – but one will never exactly know without asking them, as it is for example possible that there is some hidden reason for doing this. Of course, one can ask people in an interview what they have done in the past, but methodological research has shown that there are a number of drawbacks distorting information: People could lie (e.g. because they are ashamed of former actions); people could falsely remember the past or people could just not remember the past at all.

So although very often, different data types can be used to address the same theoretical problem, there is often a data type best suited for the problem. Additionally, there are some data (types) that cannot address specific theoretical problems at all. If one takes this statement seriously, instead of asking “Which is the best data type?”, it makes more sense to ask: What are the theoretical potentials of different data types? Can all data be used for solving a specific theoretical problem, or do different data have different advantages and disadvantages concerning this theoretical problem? How can one assess the suitability of a specific data type?

One characteristic of current research practice is that since World War II, sociological and historical research each primarily focus on one specific data type and thus have accumulated vast knowledge in the potentials and problems of using this data type, while neglecting the other.

3.1 The Sociological Focus: Research-Elicited Data

While early sociologists such as Max Weber naturally triangulated different data sources (Scheuch 1977: 36), modern sociologists have mainly based their research and focussed methodological discussion on research-elicited data (primary data) and on secondary analysis of research-elicited data: Primary data are collected by the researcher herself (e.g. observation, interviews). Secondary data were also originally collected for social science research, but are later re-analysed, maybe even for different purposes (for a discussion on secondary analysis of qualitative data, see HSR 33 (3)). While in the Anglo-Saxon tradition, ethnography is still a widespread technology, in German sociology, empirical research has been dominated by survey research ever since the 1950s. Since the late 1980s, qualitative techniques have re-surged, but here, too, oral data (i.e. interviews) are the dominating data source, in this is where most methodological knowledge has been accumulated (Currall/Towler 2003: 515-516; Baur 2005: 167; 237; 319).

At the same time, most of the knowledge on using process-generated data (process-produced data) has been lost within sociology. Process-produced data are data that are

... generated through the very processes of living, working, interacting in the societies (...) – from plain material evidence through all kinds of artifacts to the varieties of symbolic representations of ideas, activities, and events, whether drawings, tales, messages, or documents ... (Rokkan 1969: 4-5).

Process-generated data have in common with secondary data that they “were originally recorded or ‘left behind’ or collected at an earlier time by a different person from the current researcher” (Johnson/Turner 2003: 314). However, they differ from secondary data in being not originally intended for social research but being a by-product of social processes themselves. For example, newspaper articles are not written for social researchers but to inform the public about certain events.

Although in sociological methodological discussions, process-generated data are generally stated as one possible data type in overviews on social science data, usually neither their specific strengths and weaknesses (in comparison to other data types) are discussed, nor are guidelines given how to handle them in research practice (Baur 2005: 87; 280-281, 319). An example for this are discussions on mixed methods (e.g. Johnson/Turner 2003, Hunter/Brewer 2003). The little information that exists on process-generated data is not much help for research practice. For example, Currall and Towler (2003: 517) state that qualitative process-generated data are low in control, realism, access to participants, measurement precision, statistical conclusion validity, moderate in detail and high in generalizability, while quantitative process-generated data are low in control and moderate in realism, access to participants, measurement

precision, statistical conclusion validity, detail and generalizability. Unexperienced researchers can make little use of this kind of general information.

One of the possible reasons why process-generated data have so far eluded sociological methodological discussion is that sociologists tend to first classify and theorize problems (even methodological ones), before discussing them. With research-elicited data, this classification is rather easy. E.g., Behnke et al. (2006), name two dimensions of classification and discuss methodological problems along these dimensions: (a) less-structured/open data types (e.g. interviews, participant observation) and strongly structured/standardised data types (e.g. surveys, structured observation); (b) verbal data (e.g. interviews, surveys, letters) vs. visual data (e.g. observation, films).

The problem with process-generated data is that basically anything human beings have left behind can be used as a social science data source. Accordingly, process-generated data are much more complex than the usual sociological classifications of data suggest, i.e. process-generated data can be standardised and verbal (e.g. customer data bases, web logs and administrative forms), standardised and visual (e.g. marketing graphs), less-structured and verbal (e.g. documents, novels, diaries, letters) or less-structured and visual (e.g. paintings, photos, landscapes, buildings, monuments and objects). Some types of process-generated data are hard to classify, as they contain both verbal and visual or both less-structured and strongly standardised elements, e.g. GIS data, websites, films, maps, mechanical drawings and construction plans.

Despite these problems in systematically grasping process-generated data, on a methodological level, there are several reasons, why sociologists recently returned to using them. One of them is increased availability, as (at least in Germany) government agencies have started making public administrative data accessible for scientific re-analysis. Second, there are a number of new data types (especially web-data) that are also easily accessible for social science research. Thirdly, in the last 20 years, there have been great advances in data analysis procedures (especially in quantitative research) for analysing longitudinal research questions, e.g. time series analysis, event history analysis, sequence analysis, advanced panel analysis (Baur 2005). From this point of view, one could ask: Which types of data are needed to apply these procedures? Fourthly and most importantly, theoretical debates within sociology have changed (again): Many modern social theories address long-term social processes. As Scheuch (1977) pointed out: Different types of theoretical questions require different types of data for addressing them. In this case this means: For many current sociological problems, process-generated data might be better suited.

3.2 The Historical Focus: Process-Generated Data

In contrast to sociologists, historians have mainly based their research on process-produced data – with the exception of oral history. The two main reasons for this are (Baur 2005):

- 1) *Different data types are available for each epoch.* Except for contemporary history, there will be no more eye-witnesses of past events, i.e. nobody who could be enquired. The researcher, therefore, has to use data types other than interviews.
- 2) *The farther one goes back into the past, the more difficult it gets to find sources at all.* Accordingly, comprehensive procedures are required to locate historical data for a certain problem at all. It strongly depends on the epoch under review what types of data a researcher can use (Günther 2001; Küttler 1994; Meister 1997, 1999; Nünning 1995; Seiffert 1996; Opge-noorth 1997; Theuerkauf 1997).

3.3 Comparing Data Types

Strangely (and maybe with the exception of HSR), there is little methodological transfer between disciplines, although there are large overlaps in specific sub-fields of both disciplines. An example are the techniques of oral history and biographical/narrative interviews: Not only are the techniques mostly the same, but they are often used for the same kind of theoretical problems. Exchanging ideas and experienced between disciplines could lead to large synergies.

As strengths and weaknesses and the respective theoretical potentials of various research-elicited and process-generated data are rarely examined and as there is little exchange between disciplines, we today stand at the beginning of a potentially fruitful discussion. One way of reading the papers in this special issue is thus exactly about what can we learn about different data types. Table 3 gives an overview which authors addresses which data type. Some authors (e.g. Baumgarten/Grauel, Kuzmics, Schmitz et al., Young) directly compare the theoretical potentials of several data sources. Others focus on one data source, so the comparison with other data sources has to be drawn across articles. Still, as can be seen from the table, the special issue allows to compare a vast range of data sources such as interviews and surveys, public administrative data and “newer” data sources like web-data and TV commercials.

Table 3: Data Types Addressed in Articles

Data Type	Authors
Secondary Analysis of Research-Elicited Data	
Qualitative Interviews / Oral History	Freund
Surveys	Hagenah/Meulemann, Jones/Peskin, Salheiser, Schmitz et al.
Observation / Ethnography	–
Process-generated Data	
Ego-Documents and Autobiographies	Kuzmics
Popular Literature	Ernst
Public Administrational Data	Salheiser, Young
Military Records	Kuzmics, Rass
Marriage Records	Young
Genealogies	Franke, Young
Election Programs	Volkens et al.
Newspapers and Journals	Baumgarten/Grauel, Mayer
Television Commercials	Bauernschmidt
Web-Based Process-Generated Data	Baumgarten/Grauel, Schmitz et al.

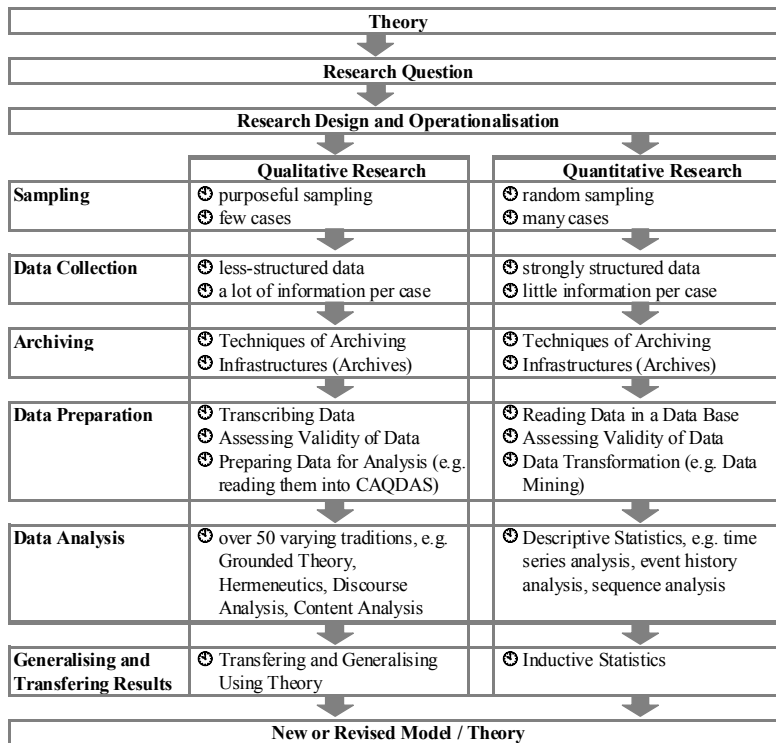
4. Stage of the Research Process Addressed in Methodological Discussion

So far, I have suggested reading the articles in this special issue either by comparing how theoretical traditions differ in addressing the problem of theory and data in longitudinal research, or by comparing which type of data are suited for grasping for which types of theoretical problems. In both cases, I have generally talked about “theory” and “data”. If one regards research practice, a third way of tackling the problem of theory and data comes to mind:

In research practice, we have to tackle a number of problems, which we usually handle either sequentially (a model preferred in quantitative research) or parallelly or iteratively (models preferred in qualitative research). Idealtypically, one could say that researchers start with a specific research question, which they address from a specific theoretical standpoint. Depending on theoretical tradition, researchers either want to test the theory, build or modify new theories. However, as stated above, the starting point is always an existing theoretical perspective. As figure 1 illustrates, using this theoretical perspective, researchers specify their research question, decide on research design and on operationalization. Regardless, if they want to test existing or build new

theories, they have to define a population and select cases, collect data, prepare them for analysis, analyse them and decide if and how results can be generalised or transferred to other contexts. As longitudinal data are usually generated over long time-spans, they also have to be archived and re-accessed. For each of these stages, researchers have to decide, if their research is valid. At the end of this process, there is usually a new or a modified theory or model.

Figure 1: Traditional Model of Linking Theory and Data During the Research Process



From the point of view of the research process, one can thus ask for each stage of the research process, how theory and data are related within this stage, i.e.: (How) does theory influences, how this specific stage of the research process is handled? Do different theories address different stages or point out different problems within each stage? Vice versa, how does the methodological handling of data within this stage affect theory building, theory testing and overall validity?

Comparing (sociological) methodological discussion on research-elicited data and (historical) methodological discussion on process-produced data, the first thing that catches the eye is that they focus on different stages of the research process and handle each stage in different ways. For example, sociology usually neglects the issue of archiving and preparing data, while there is a lot of discussion on sampling and interpreting/analysing data.

In history, the situation seems vice versa, putting a strong focus on data preparation: Before starting the actual interpretation process, researchers first need to examine sources' authenticity, prepare them and criticise them. Historians therefore generally encounter any type of data with mistrust: Each source blurs the view to what people actually thought or did, i. e. data and facts do not simply exist, but are rather constructed by people and need to be interpreted correspondingly. In order to be able to assess the quality and explanatory power of sources, researchers have to look in-depth into the sources, have to gather context information and have to question it critically over and again. Still, despite all efforts of decoding and deconstruction, researchers will never completely overcome the data's specific perspective (Günther 2001; Küttler 1994; Meister 1997, 1999; Nünning 1995; Seiffert 1996; Opgenoorth 1997; Theuerkauf 1997). This careful scrutinising of data is even today one of the strengths of historical methods in comparison to modern sociological methods (Baur 2008b).

If one applies the general division of research-elicited data (i.e. surveys, interviews and observation) and process-generated data on discussion the research process and discusses its implication for theory building on a general level, one can say that process-generated data have several advantages in comparison to research-elicited data: Process-generated data are non-reactive. They can be used, if other means of data collection are not applicable, for example, if infrastructure for large-scale surveys does not exist (which is the case in many countries of transition), if response-rates in surveys are expected to be too low, if researchers might not get access to interview partners or if the social phenomenon of interest is not observable (e.g. when analysing past events or hidden populations).

At the same time, process-generated data face several problems, the severest being that, from a statistical point of view, they are usually *biased*. Bias is already likely to occur during data production, as process-generated data are not generated for scientific but for practical purposes (Baur 2004). *How* they are they biased depends on (1) the particular purpose, (2) format and (3) institutional embeddedness of the pertinent data type. In addition, all three elements may change over time (Baur/Lahusen 2005). Finally, the original data might decay or will be destroyed deliberately. The process of data selection is biased, too, as humans have to actively want to preserve data available for later use (Baur 2004, Baur/Lahusen 2005).

These classifications are rather general. As we know from sociological methodological research, what is needed, is a methodological discussion on the advantages and disadvantages of how specific data types, specific theoretical questions and specific phases of the research process are intertwined and if one can draw general conclusions from this.

Table 4: Stage of Research Design / Methodological Problem

Data Type	Authors
Sampling / Data Selection	
Defining the Population	Mayer, Rass
Hidden Populations	Franke
Accessing Data / Gatekeepers	Baumgarten/Grauel, Bauernschmidt
Selecting Materials	Baumgarten/Grauel, Ernst, Mayer, Rass, Volkens et al., Young
Assessing Production Bias	Baumgarten/Grauel, Franke, Freund, Rass, Salheiser, Young
Assessing Selection Bias (Archiving, Data and Information Loss)	Baumgarten/Grauel, Franke, Bauernschmidt, Freund, Rass
Interpreting Data / Data Analysis	
Operationalizing Theoretical Problems	Kuzmics, Mayer, Schmitz et al., Volkens et al.
Building Theories	Baumgarten/Grauel, Ernst, Freund
Testing Theories	Jones/Peskin, Schmitz et al.
Using Context Information	Hagenah/Meulemann, Jones/Peskin, Young
Content Analysis	Bauernschmidt, Ernst, Mayer, Volkens et al.
Assessing Validity	
Mixing, Triangulating and Integrating Data	Kuzmics, Salheiser, Volkens et al. Baumgarten/Grauel, Franke, Kuzmics, Schmitz et al., Young

The contributors to this special issue try to tackle these questions. Therefore, a third way of reading the papers is by comparing what authors say on specific phases of the research process. All authors have in common that they are interested in questions of historical sociology and in social processes and that they cannot draw on primary data for their analysis. Thus, not of the articles discusses how to organize data collection. Although some authors (Franke, Freund, Salheiser, Volkens et al.) also touch issues of data preparation and archiving, these questions are marginal to this special issue. Instead, HSR 34.3 will focus on advances in data preparation. As table 4 illustrates, the phases most articles in this special issue focus on, are sampling and data selection, data interpretation and assessing validity.

For each phase, one can identify methodological sub-themes: During *sampling*, researchers have to define the population, which is especially a problem for hidden populations. As data are usually buried in archives, researchers have to identify the right archive and get access to the data. Once having gained access to data, usually materials have to be selected. Finally, researchers have to assess how data were biased during data production and due to selection processes (such as data loss) and how this affects interpretation and theory building.

Especially with less-structured data, *data analysis* poses another bundle of unsolved questions: First of all, how can process-theories be operationalized for empirical research? Second, how can data be interpreted in order to build or test process-oriented theories? Thirdly, how can context information be used in order to either assess the validity of interpretation or to further theory building? Various authors make different suggestions for tackling this problem. Interestingly, several authors suggest applying or modifying content analysis during this process.

Finally, several authors elaborate procedures for assessing *validity* either of the whole research process, for specific phases of the research process or for the whole research process. Specifically, some authors try to apply models of mixing and triangulating data to process-generated data. All in all, one can conclude from these articles that theory and data are closely entwined during all phases of the research process and that there are still a lot of open questions.

5. References

- Baur, Nina (2004): Wo liegen die Grenzen quantitativer Längsschnittsanalysen? Series: Bamberger Beiträge zur empirischen Sozialforschung 23. Bamberg.
- Baur, Nina (2005): Verlaufsmusteranalyse. Methodologische Konsequenzen der Zeitlichkeit sozialen Handelns. Wiesbaden: VS-Verlag.
- Baur, Nina (2008a): Taking Perspectivity Seriously. In: Historical Social Research 33 (4). 191-213.
- Baur, Nina (2008b): Was kann die Soziologie methodisch von der Geschichtswissenschaft lernen? In: Historical Social Research 33 (3). 217-248.
- Baur, Nina/Lahusen, Christian (2005): Sampling Process-Generated Data in order to Trace Social Change. In: van Dijkum, Cor/Blasius, Jörg/Durand, C. (Eds.): Recent Developments and Applications in Social Research Methodology. Opladen: Verlag Barbara Budrich (e-book).
- Behnke, Joachim/Baur, Nina/Behnke, Nathalie (2006): Empirische Methoden der Politikwissenschaft. Paderborn: Ferdinand Schöningh.
- Currall, S. C./Towler, A. J. (2003): Research Methods in Management and Organizational Research. In: Tashakkori, A./Teddlie, C. (Eds.): Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks et al.: Sage.513-526.
- Esping-Andersen, Gøsta (1990): The Three Worlds of Welfare Capitalism. Cambridge/Oxford: Polity/Blackwell.
- Glaser, Barney/Strauss, Anselm (1975): Awareness of Dying. Chicago: Aldine.

- Günther, Rosmarie (2001): Einführung in das Studium der Alten Geschichte. Paderborn et. al.: Schöningh.
- Hunter, A./Brewer, J. (2003): Multimethod Research in Sociology. In: Tashakkori, A. & Teddlie, C. (Eds.): Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks et al.: Sage. 577-594.
- Johnson, R. B./Turner, L. A. (2003): Data Collection Strategies in Mixed Methods Research. In: Tashakkori, A./Teddlie, C. (Eds.): Handbook of Mixed Methods in Social and Behavioral Research. Thousand Oaks et al.: Sage. 297-320.
- Kalthoff, Herbert (2008): Zur Dialektik von qualitativer Forschung und soziologischer Theoriebildung. In: Kalthoff, Herbert/ Hirschauer, Stefan/ Lindemann, Gesa (Eds.): Theoretische Empirie. Frankfurt a.M.: Suhrkamp. 8-34.
- Kalthoff, Herbert/Hirschauer, Stefan/Lindemann, Gesa (Eds.) (2008): Theoretische Empirie. Zur Relevanz qualitativer Sozialforschung. Frankfurt a.M.: Suhrkamp.
- Knoblauch, Hubert (2008): Sinn und Subjektivität in der qualitativen Sozialforschung. In: Kalthoff, Herbert/Hirschauer, Stefan/Lindemann, Gesa (Eds.): Theoretische Empirie. Frankfurt a.M.: Suhrkamp. 210-233.
- Küttler, Wolfgang (1994): Geschichtstheorie und -methodologie der DDR. In: Zeitschrift für Geschichtswissenschaft 42 (1). 8-20.
- Lindemann, Gesa (2008): Theoriekonstruktion und empirische Forschung. In: Kalthoff, Herbert/Hirschauer, Stefan/Lindemann, Gesa (Eds.): Theoretische Empirie. Frankfurt a.M.: Suhrkamp. 165-187.
- Meister, Klaus (1997): Einführung in die Interpretation historischer Quellen. Schwerpunkt: Antike. Volume 1: Griechenland. Paderborn et al.: Schöningh.
- Meister, Klaus (1999): Einführung in die Interpretation historischer Quellen. Schwerpunkt: Antike. Volume 2: Rom. Paderborn et al.: Schöningh.
- Nünning, Vera (1995): Uni-Training Geschichtswissenschaft. Stuttgart/ Dresden: Klett.
- Opgenoorth, Ernst (1997): Einführung in das Studium der neueren Geschichte. Paderborn et. al.: Schöningh.
- Rokkan, Stein (1969): Comparative Cross-National Research: The Context of Current Efforts. In: Merritt, Richard L./Rokkan, Stein (Eds.) (1963): Comparing Nations. New Haven; London: Yale University Press. 3-25.
- Scheuch, Erwin K. (1977): Die wechselnde Datenbasis der Soziologie. Zur Interaktion zwischen Theorie und Empirie. Reprinted in: Historical Social Research, Supplement 18 (2006). 24-46.
- Seiffert, Helmut (1996): Einführung in die Wissenschaftstheorie. Zweiter Band: Geisteswissenschaftliche Methoden. München: C. H. Beck.
- Theuerkauf, Gerhard (1997): Die Interpretation historischer Quellen. Schwerpunkt: Mittelalter. Paderborn et. al.: Schöningh.
- Wehler, Hans-Ulrich (1980): Historische Sozialwissenschaft und Geschichtsschreibung. Göttingen: Vandenhoeck & Ruprecht.