Z Erziehungswiss (2011) 14:301–313
DOI 10.1007/s11618-011-0191-6

ZfE

19 Data protection issues in the National Educational Panel Study

Sonja Meixner · David Schiller · Jutta von Maurice · Henriette Engelhardt-Wölfler

Abstract: In an information- and knowledge-based society, data protection plays a significant role. Basically, it has to ensure the right to informational self-determination codified in the individual's right to decide whether to disclose or not disclose his or her personal data. Recent decades have seen a strong growth in the awareness of data protection issues in the social sciences. The German National Educational Panel Study was established to collect survey data on educational processes and competence development for the scientific community. Its complex multicohort sequence design harbors several challenges for data protection: The legal regulations, the longitudinal design, the different populations under study, the varying collection modes and the sampling procedure all need to be considered from the perspective of data protection when collecting, processing, and disseminating data. Appropriate procedures and clear structures are essential. These can be developed only in a close cooperation between social scientists and data protection experts.

Keywords: Data protection · Social sciences · Education · Panel study

Datenschutz im Nationalen Bildungspanel

Zusammenfassung: In einer informations- und wissensbasierten Gesellschaft kommt dem Datenschutz eine bedeutende Rolle zu. Das Recht auf informationelle Selbstbestimmung, welches es dem Individuum erlaubt, über die Preisgabe seiner persönlichen Daten selbst zu entscheiden, ist grundsätzlich zu schützen. In den letzten Jahren hat das Bewusstsein über die Bedeutung des Datenschutzes in den Sozialwissenschaften stark zugenommen. Das Nationale Bildungspanel wurde ins Leben gerufen, um Daten über Bildungsverläufe und Kompetenzentwicklung für die wissenschaftliche Nutzung zu erheben. Das komplexe Multi-Kohorten-Sequenz-Design birgt mit Blick

Dipl.-Soz. S. Meixner · D. Schiller, M.A. · Dr. J. von Maurice
National Educational Panel Study, University of Bamberg, 96045 Bamberg, Germany
e-mail: sonja.meixner@uni-bamberg.de

D. Schiller, M.A.
e-mail: david.schiller@uni-bamberg.de

Dr. J. von Maurice
e-mail: jutta.von-maurice@uni-bamberg.de

Prof. Dr. H. Engelhardt-Wölfler (✉)
Chair of Population Studies, University of Bamberg, 96045 Bamberg, Germany
e-mail: henriette.engelhardt-woelfler@uni-bamberg.de

auf datenschutzrechtliche Fragen zahlreiche Herausforderungen: Die rechtlichen Grundlagen, das Längsschnittdesign, die unterschiedlichen Untersuchungspopulationen sowie die vielfältigen Methoden der Datenerhebung und Stichprobenziehung müssen aus datenschutzrechtlicher Perspektive bei der Datenerhebung, Datenverarbeitung und Datenweitergabe betrachtet werden. Geeignete Prozeduren und klare Strukturen sind hierbei von zentraler Bedeutung. Diese können nur in enger Zusammenarbeit von Sozialwissenschaftlern und Datenschutzexperten entwickelt werden.

Schlüsselwörter: Datenschutz · Sozialwissenschaften · Bildung · Panelstudie

19.1 Introduction

Data protection is one of the most important acceptance factors for the development of modern information- and knowledge-based societies (Bizer 2007). A survey by the Allensbach Institute for Public Opinion Research in 2009, however, indicates that more than 60% of the German population worries about insufficient data protection; more than one-half of the respondents (52%) even say that they have become more cautious when asked to give data about themselves (Institut für Demoskopie Allensbach 2009). At the same time, more and more data is being produced, stored, and processed as a result of new technical advances. In the course of the rapidly expanding bulk of data, we hear about misuse of data, data leaks, identity theft, or illegal video surveillance in the media almost every day. Newspaper articles or broadcasts on these topics have become part of our daily lives. Although these incidents (e.g., violation of privacy by spying on employee data)[1] do not extend into the field of scientific research, they reveal the importance of data protection in all areas of modern life.

For scientific (empirical) research, the collection and use of data is essential. Therefore, data protection issues in data collection and data use have to be a major priority in the research projects planned and conducted by all scientific disciplines. Consequently, advances in scientific research go hand in hand with advances in data protection. However, decisive progress in this area requires a detailed discussion of problems and their possible solutions as well as a close cooperation between scientific researchers and data protection experts. Recent decades have seen an ongoing discussion on the needs of data collection and data use in scientific contexts and on data protection issues. This process has led to, for example, modified Data Protection Acts and court decisions specifying data protection regulations. There is also a growing awareness of these issues in the social sciences, as can be seen in the publication of data protection concepts for social research projects (see Frick et al. 2010) and continuous research on statistical disclosure control (Hundepool et al. 2010; Ichim and Franconi 2010; Shlomo et al. 2010).

The National Educational Panel Study (NEPS) has been set up to collect longitudinal data on educational processes and competence development. Its research goals and the complex multicohort sequence design harbor several challenges for data protection. This chapter outlines these data protection challenges and corresponding procedural-organizational stipulations within the NEPS. It describes the kinds of data in social research, the consequences of the multicohort sequence design for data protection issues, and the legal regulations. This then serves as a background to focus on the implementation of data protection in the areas of data collection, data preparation, and data dissemination. Data

protection in the sense of protecting data from getting lost, for example, by making copies of it or storing it in a secure environment, is not the focus of attention here.

19.2 Survey data in the social sciences

The legal foundation for data protection in Germany is the German Federal Data Protection Act (Bundesdatenschutzgesetz, BDSG) as well as the Data Protection Acts in each of the 16 federal states (Landesdatenschutzgesetze). They aim to protect the individual's personal rights. As Section 1 (1) BDSG states: "Zweck dieses Gesetzes ist es, den Einzelnen davor zu schützen, dass er durch den Umgang mit seinen personenbezogenen Daten in seinem Persönlichkeitsrecht beeinträchtigt wird."[2] This right to privacy also includes the right to informational self-determination. It derives from Article 2 (1) of the Basic Constitutional Law of the Federal Republic of Germany (Grundgesetz, GG): "Jeder hat das Recht auf die freie Entfaltung seiner Persönlichkeit, soweit er nicht die Rechte anderer verletzt und nicht gegen die verfassungsmäßige Ordnung oder das Sittengesetz verstößt"[3] in conjunction with Article 1 (1) GG "Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller staatlichen Gewalt."[4] The right to informational self-determination guarantees the protection of the individual from unregulated disclosure and utilization of personal data. An adjudication by the German Federal Constitutional Court states that the individual should always decide these issues personally (BVerfGE, Entscheidungen des Bundesverfassungsgerichts, 65, 1).

According to the legal definition in Section 3 (1) BDSG, *personal data* is defined as "Einzelangaben über persönliche oder sachliche Verhältnisse einer bestimmten oder bestimmbaren natürlichen Person (Betroffener)."[5] On the one hand, the scope of protection accordingly covers personal data allowing a direct identification of a natural person (e.g., via name, personal picture, address, phone number, or social insurance number). On the other hand, it also refers to person-related data. That kind of data does not allow a clear or immediate identification of the respondent via "direct identifiers" but via additional information derived from other data sources (e.g., information given by friends, the media, etc.) and via the combination of several single pieces of information (e.g., combination of occupation, place of residence, and migration background; Häder 2009; Metschke and Wellbrock 2000).

In the field of social science, the units of analysis are individuals. Therefore, gathering data about individuals is a fundamental need for social research. It is also the basis for statistical methodology that successively develops new statistical models designed to explain social phenomena, changes in society, or human development. However, social scientists are not interested in specific individuals but in representatives of populations under study. The aim is not to assess individual characteristics, but to obtain generalizable results. As a consequence, statistical analyses in social science do not require the identity of single individuals, and there is no need to work with personal data. It is sufficient to work with survey data.

Survey data is a dataset belonging to individuals who have participated in a survey. Main characteristics of survey data are, first, that each individual in the dataset is defined by a unique code such as an ID (the code itself should not allow a direct connection to

an individual; instead, it should be a real alias). Second, the entity of data in the dataset belonging to a single individual should not allow any reidentification of the person. Methods of pseudonymization and anonymization are necessary to meet this requirement (see Sect. 19.5). In general, survey data needs to be of high quality if one is to obtain significant and reliable results. The basis for high data quality is especially a reasonable deduction of questions, a good operationalization of constructs, a well-constructed sampling design, an adequate data collection process, and a high and representative response rate.

Survey data are essential for making substantial progress in social research. The freedom of science is guaranteed by Article 5 (3) cl. 1 GG: "Kunst und Wissenschaft, Forschung und Lehre sind frei."[6] As Metschke and Wellbrock (2000) point out, this freedom of science may collide with general personal rights in data collection. The challenge for social science is to find an acceptable compromise between realizing the freedom of science and guaranteeing general personal rights (Häder 2009; Metschke and Wellbrock 2000). Meeting both scientific requirements and data protection regulations is the general guideline for all activities within the NEPS.

19.3 Data protection challenges in the complex multicohort sequence design

The NEPS is one of the largest longitudinal studies ever started in the field of education. More than 60,000 target persons of different age groups will be questioned and tested regularly (see Chap. 1, this volume). Its multicohort sequence design is quite challenging, not only for the scientific researchers developing the methods and instruments but also for the project coordination staff of the NEPS and the data-collecting institutes who are implementing the data collection procedures. The ways of accessing respondents, the recruiting processes for different target groups, and the processes of field work have to be specified. Data protection has highest priority in all these aspects. The interplay between the main legal regulations, the implications of the longitudinal design, the different ages of the populations under study, the varying data collection modes (each connected with different procedures to gain consent to participate), and the hierarchical structure of data have crucial implications for data protection:

- Before setting up a data collection process within the NEPS in a specified substudy, it is necessary to clarify the legal regulations. These differ depending on the context of the data collection. For example, recruiting students for the NEPS in schools requires different processes compared to recruiting participants via register-based data (see Sect. 19.4).
- One important characteristic of the NEPS is its longitudinal design. Whereas respondents are contacted only once in cross-sectional studies, the NEPS follows all target persons for years. To reapproach our target persons in subsequent panel waves, the NEPS needs to ask the respondents for contact data (i.e., name, address, e-mail address, telephone number). For data protection reasons, the NEPS has decided not to store the contact information in the NEPS coordination center in Bamberg, but to store it—strictly separated from the survey data—at one of the data-collecting institutes (see Sect. 19.5).

- Managing different cohorts from newborns to adults is a big and difficult task from a data protection perspective. First of all, we have to clarify responsibilities for giving consent to participate in different populations. Whereas asking adults for their participation is quite uncomplicated, the situation becomes more complex when minors are included in a sample. Here the interplay between parental consent and the child's consent has to be clarified taking the age of the child into account (see Sect. 19.5).
- Furthermore, a variety of data collection modes are used, ranging from written questionnaires and competence tests, across interviews in a face-to-face or telephone mode, to online surveys. The way of asking for consent needs to be adjusted to the way of contacting the respondent. Of course, this also needs to be taken into account when providing participants with further information about current issues in the study at a later date (see Sect. 19.5).
- In addition, the reference to institutions such as schools or Kindergartens in the sampling procedure in some NEPS cohorts has implications for data protection concepts. These institutions themselves are worthy units of protection. When generating, for example, survey data for the scientific community, the aspect "additional information" (e.g., participant X attended school Y or Kindergarten Y) plays an important role and needs to be considered (see Sect. 19.6).

Altogether, many aspects need to be discussed when handling data collection in the multicohort sequence design of the NEPS. Certainly, when conflicts emerge between data protection issues and scientific requirements, the staff of the NEPS is highly committed to data protection regulations and cooperates closely with data protection experts in developing good solutions. Such a commitment also strengthens the respondents' confidence that needs to remain positive over the course of such a large-scale project.

19.4 Legal regulations

The NEPS has to consider various legal regulations for data collection, data handling, and data dissemination. The Bavarian Data Protection Act (Bayerisches Datenschutzgesetz, BayDSG) is the designated law for the NEPS coordination center of located at the University of Bamberg as a public institution of the Federal State of Bavaria. With the Bavarian Data Protection Act as the guiding framework for the NEPS coordination center, the University Data Protection Officer and the Bavarian State Commissioner for Data Protection and his team support and accompany the NEPS in data protection issues.

The legal basis of data collection has to be examined in more detail. In general, NEPS data is collected by professional data-collecting institutes (see Chap. 1, this volume). These institutes are bound by the German Federal Data Protection Act, and the Commissioners for Data Protection in the federal states in which the institutes are registered are responsible for controlling their operations (independent of the individual study commissioned). When starting the collaboration between NEPS and these institutes, the schedule of responsibilities and the compliance with data protection issues in data collection and data transfer had to be regulated carefully. A special case of jurisdiction in Germany is data collection in the school context (e.g., data collection in the 5th- and 9th-grade

starting cohorts). In each of the 16 federal states, the particular Ministry of Education inspects the instruments, materials (e.g., information given about data protection to the participants), and the data collection procedures with regard to their content and data protection aspects. Here, priority is given to the respective Education Act (Schulgesetz). In many cases, however, the Ministries of Education refer to the Data Protection Act of their particular federal state or the German Federal Data Protection Act. When engaged in the verification process required for data collection in the school context, the Ministries of Education in the 16 federal states are in close contact with the NEPS coordination center in Bamberg. Negotiations focus on finding appropriate solutions for all 16 federal states in order to avoid as far as possible any distortions due to federal-state-specific adjustments to instruments, materials, and procedures.

As a result of the processes described, the NEPS team is continuously optimizing instruments, materials, and procedures in compliance with data security aspects. Many appropriate solutions have been found for difficult data protection issues (e.g., in the area of context questions or analyses of the underlying population). Despite the existence of a general German Federal Data Protection Act and the accompanying 16 different Data Protection Acts of the federal states, the core elements of these laws are quite similar. However, slight differences between the formulations of the laws in the 16 federal states and the room for interpretation in each legal situation demand very exact examinations of data protection issues and an ongoing exchange between all persons in charge.

Because of the dominant role of the Bavarian Data Protection Act for the NEPS coordination center in Bamberg, and in order to restrict the following discussion to central aspects (leaving federal-state-specific aspects unconsidered), we shall refer to this law when explaining the collection, processing, and utilization of the NEPS data.

19.5 Data collection process

The participants in the NEPS are selected through random samples that differ between the six starting cohorts (see Chap. 4, this volume). In Bavaria, for example, data collection, processing, and utilization are regulated by Article 15 (1) BayDSG. This states that collecting, processing, and utilizing personal data is only allowed if a law or a different legal regulation allows or provides it, or if concerned persons agree to it. In the case of the NEPS, there is no law obliging people to participate in the study; rather, it is every single person's own and free decision. The second part of paragraph 1 therefore legitimizes the survey process. Peoples' consent to participate is needed; collecting, processing, and using personal data against a person's will is not permitted.

Freedom of decision also means that everybody can determine the way in which and the extent to which their personal are processed (Metschke and Wellbrock 2000). It is therefore essential for people to be able to estimate the full consequences of their participation in the NEPS before giving their consent.

Article 15 (2) BayDSG stipulates which information has to be given when asking people for their consent to participate in a survey. First, they need to be informed about the purpose of the data collection, the data processing, and the data utilization. Second, the receivers of their personal data have to be named. Third, the possibility of refusing

consent has to be indicated explicitly. And, last but not least, people need to be informed about the consequences of refusing their consent—because participation in the NEPS is voluntary, nobody needs to fear any disadvantages by refusing. In addition, it is statutory for people to also be informed about their right to withdraw their given consent at any time. Basically, on the one hand, information about the study needs to be adequate and sufficient enough to ensure a valid consent. On the other hand, every individual should be able to understand it regardless of their education background. Realizing both requirements is quite a balancing act for the NEPS.

According to Article 15 (7) BayDSG, there is a set of data requiring special treatment when collecting, processing, or using it. Data belonging to this set addresses "die rassische und ethnische Herkunft, politische Meinung, religiöse oder philosophische Überzeugung oder die Gewerkschaftszugehörigkeit … sowie … Daten über Gesundheit oder Sexualleben."[7] Disclosure of such data could have especially harmful results for individuals. According to this regulation, the NEPS is allowed to collect this kind of data only if the participants agree, and if their agreement refers explicitly to this kind of data. To answer the most current and important research questions on education and competence development in Germany, it is absolutely essential to collect data on peoples' migration background, the languages they speak—both indicators of "racial and ethnic origin"—and data about their religious life (see Chap. 8, this volume). Taken together, such sensitive data can be collected and used for scientific research; however, great care is needed.

Generally, the information given to the participants is the basic element for their consent, and this is absolutely obligatory for researchers. Insofar, "informed consent" frames the data collection, data processing, and data utilization process. However, panel studies are not static but develop over years. New research topics could evolve that have not been covered by the original consent. In that case, researchers would have to ask for consent again later on, should that be possible. Another option would be to formulate the declarations of consent in a broader way right from the start when recruiting participants at the beginning of the study—this procedure is more compatible with scientific working methods. In the end, it is the peoples' individual and free decision whether they accept the more broadly formulated declaration of consent or not (Metschke and Wellbrock 2000). One big advantage of the panel design of our study is that we always stay in close contact with our participants. Therefore, we can easily inform them about a new main focus or about new developments in questioning should that be required.

Another important data protection aspect is regulated by Article 15 (3) cl. 1 BayDSG. This states that consent basically has to be given in written form unless special circumstances require another form. The written form protects people against a too hasty or thoughtless consent. On the basis of the given information, they should first think about the consequences of their consent. At the same time, the written form is always evidence for legitimate data collecting, processing, and utilization, and it ensures the transparency of the data-collecting process. However, a written consent also leads to relevant problems in the field of social research. The nonresponse rate might increase in certain social groups—for example, people who fear fine print—and endanger the representativeness of the sample. Apart from that, telephone interviews play an increasingly important role in the social sciences for cost reasons, but asking respondents for written consent is often either impossible or very difficult. Researchers often do not have peoples' addresses,

or first need to collect the current one. The Data Protection Act allows some flexibility through the expression "soweit nicht wegen besonderer Umstände eine andere Form [der Einwilligung] angemessen ist."[8] Thus, an explicit oral consent may replace the written one in some cases; particularly in telephone interviews (Metschke and Wellbrock 2000). In the NEPS, we ask our target persons to sign a declaration of consent whenever possible.

As the NEPS is analyzing education across the entire life course and competence development from birth to adult life, our target persons are of different ages and many of them are minors (under 18 years). Basically, parents are responsible for their minor children. In the NEPS, we always ask the parents to permit their minor child's participation in the study (see also Brocks 2009). Of course, we also need to respect the child's will, and we need to accept his or her decision not to participate despite the parents' written consent—participation in our study is also voluntary for the children involved. Apart from that, Germany's Basic Constitutional Law grants every child the same basic rights as an adult, and consequently also the right to informational self-determination. In order to fulfill that legal condition, we also ask each minor child to give explicit consent. Generally, this consent is only valid and effective if the individual has the ability to form a rational judgment about the issue, and this also includes understanding the consequences of the consent. Unfortunately, the different laws connected to data protection do not define an age limit for this. Ideally, one should check each potential participant's ability to make a rational judgment. Of course, this is not possible for the large number of persons in our sample. Moreover, there are no objective criteria to support such a procedure. For these reasons, a general guideline is favored. For example, according to German Criminal Law, minor children are assumed to be of age at 14 years. Thus, we only ask children aged 14 years or older to give written consent, and assume that, at this age, they are able to foresee the consequences of their participation in the NEPS.

Because the NEPS consists of several waves, we need the participants' contact information so that we can reach them and question them again some months or years later. Article 23 (3) BayDSG provides a strict separation that allows a clear reidentification of participants' data such as name, address, telephone number, e-mail address, and the data disclosed during the survey. In other words, direct identifiers, which would make it easy to reidentify a natural person, have to be separated from the survey data as soon as possible. Accordingly, the data collection process in the NEPS has been structured in a way that only the commissioned data-collecting institutes receive these contact data; they administrate the data collection process and are in close contact with the respondents. The coordination center in Bamberg receives only the survey data (see Sect. 19.2).

In conclusion, many data protection aspects need to be considered during the process of recruiting respondents for the NEPS who will be questioned several times. One particular concern is to ensure that all the above-mentioned aspects are transferred into clear procedures that are implemented in all information letters or forms used in the data collection process.

19.6 Data preparation and data dissemination

After completing the data collection in each wave, the data-collecting institutes send the data in a pseudonymized form to the NEPS coordination center. Section 3 (6a) BDSG stipulates that pseudonymization or "aliasing" is "das Ersetzen des Namens und anderer Identifikationsmerkmale durch ein Kennzeichen zu dem Zweck, die Bestimmung des Betroffenen auszuschließen oder wesentlich zu erschweren."[9] According to this, the code should be constructed in such a way that nobody will be able to reidentify a participant by the code. At the same time, a clear identification code per participant is needed because it is essential for a panel study such as the NEPS to be able to match participants' data from one wave to that from another wave. Taken together, no personal data will be delivered to the NEPS coordination center in Bamberg. Cutting of the "direct identifiers" within the data-collecting institutes and delivering only pseudonymized data to the NEPS already fulfils a first important step toward anonymization. As a result, the NEPS coordination center deals only with survey data.

When the pseudonymized data arrives in the NEPS coordination center, the codes used by the data-collecting institutes are replaced by new ones; the new codes are the ones given to the scientific community. Basically, codes are replaced only for data protection purposes. After that step, data anonymization—one of the most important legal requirements for data dissemination—data editing, and data documentation can start.

Even survey data has to be checked for its disclosure risk. The anonymization concept applied to the NEPS data follows two principles: First, disclosing respondents should be impossible. Second, a high utility of the data should be maintained. Different expressions are used to describe the levels of anonymization: Formal anonymization is achieved by dropping direct identifiers. Absolute anonymization lowers the disclosure risk to zero. However, this simultaneously reduces the data utility to zero as well. Therefore, the most important level of anonymization is factually anonymous data.

The German legislative level recognized the need for factually anonymous data when carrying out its first census in the 1980s. Based on that experience, the Federal Constitutional Court first proclaimed that collected data should be anonymized at the earliest stage possible, and second, that (factually) anonymized data meets the requirements of the constitution. Section 3 (6) BDSG defines "rendering anonymous" as "das Verändern personenbezogener Daten derart, dass die Einzelangaben über persönliche oder sachliche Verhältnisse nicht mehr oder nur mit einem unverhältnismäßig großen Aufwand an Zeit, Kosten und Arbeitskraft einer bestimmten oder bestimmbaren natürlichen Person zugeordnet werden können."[10] This definition is oriented toward the principle of comparativeness; it meets not only the individuals' right to informational self-determination but also ensures data utility for scientific research. Techniques of creating (factually) anonymized data are summarized under the term statistical disclosure control (see bullet point "*Statistical data protection*"). In the field of data dissemination, anonymization techniques are only one part of the entire data protection concept.

Within the NEPS, setting up a comfortable and secure data access is guided by a portfolio approach (Lane et al. 2008). Five different approaches are combined to protect the collected data and the respondents' identity. Strategies for data protection accordingly

include organizational, legal, statistical, educational and technical data protection (see also Chap. 20, this volume).

- *Organizational data protection.* According to the NEPS mission, the data it collects should only be available for scientific use. Commercial institutions or private persons should not gain access. Prior to allowing access to the data or transmitting it to somebody, the staff of the NEPS User Service Department at the coordination center in Bamberg screen the potential data user's status and check whether he or she is connected to a university or a noncommercial scientific research organization. Access to the data is conditional on the user belonging to the scientific community. Furthermore, the user is requested to present his or her research project to the NEPS staff in order to confirm the scientific interest. This procedure enables the staff to provide the individual researcher with only the necessary data for his or her specific project.
- *Legal data protection.* In Germany, working with research data is subject to different legal rules legislated by the Federal Republic of Germany or the 16 federal states. Our principal task is to assure compliance with the legal regulations when giving researchers access to our data. The data users therefore are provided with data protection and data security information when asking for data access. In addition, they have to sign a contract regulating important aspects of these issues. The most essential one demands a commitment from the data users to observe data secrecy. Data confidentiality is regulated in Section 5 BDSG as "Den bei der Datenverarbeitung beschäftigten Personen ist untersagt, personenbezogene Daten unbefugt zu erheben, zu verarbeiten oder zu nutzen (Datengeheimnis). … Das Datengeheimnis besteht auch nach Beendigung ihrer Tätigkeit fort."[11]
- *Statistical data protection.* Using techniques of statistical data protection means modifying the data in a way that guarantees the respondents' privacy. The aim is to create factually anonymous data that guarantees privacy while simultaneously offering a high level of data utility. In the field of social science, research on statistical data protection is an ongoing project. The methods can be subsumed under the expression statistical disclosure control. A wide range of modifications can be used to alter the data by, for example, aggregating the original data (e.g., no detailed occupation data but only the branch of the economy), adding noise to it (e.g., modifying the values of variables by defined methods), or synthesizing the original data (Hundepool et al. 2010; Rubin 1993). The results of analyses can also be altered by techniques of statistical disclosure control in order to protect the respondents' privacy. The required methods mainly depend on the technical form of data access (see bullet point "*Technical data protection*") and on the disclosure risks of the data. Generating factually anonymous data in the context of a longitudinal survey is much more challenging than working with cross-sectional data. To evaluate the disclosure risk in a dataset, it is necessary to check all variables and combinations of variables. Furthermore, the possibilities of accessing additional information have to be taken into account. In longitudinal surveys, data is collected in multiple waves and data from new waves is merged with data from the existing ones. Therefore, no final check for disclosure risk can be performed, because the kind of data to be collected in the following waves is still unknown. Due to the increasing number of longitudinal surveys, there is a grow-

ing need for more research on this topic. Finding appropriate methods of statistical disclosure control for longitudinal surveys is absolutely essential.

- *Educational data protection.* Good research principally depends on good education. Accordingly, the NEPS staff offers a special training program to data users. The program includes lessons on the complex panel design of the NEPS and the resulting data structure as well as lessons on data protection and data security. One main objective of the program is to provide researchers with sufficient information about secure scientific research.
- *Technical data protection.* The data collected within the NEPS is digital data. Technical data protection in the form of hard- and software solutions is therefore essential. Two different fields of data protection can be distinguished here: data storage and data dissemination. Concerning data storage, the staff is building a database system based on an autonomous server structure in which data is protected against both data loss and attacks from outside. In matters of data dissemination, the NEPS wants to offer a comfortable data access to researchers (see Chap. 20, this volume). Depending on the form of data access, appropriate technical data protection methods need to be installed—the level of technical data protection principally corresponds to the level of statistical data protection. When analyzing high-detail data, researchers need to work within the NEPS building. There is a workroom equipped with special computers that, for example, do not allow data to be copied and that are not connected to the Internet. Thus, a more detailed version of the data required can be offered to researchers there. Another form of data access is via a secure remote access (NEPS Secure Data Access). Scientists connect their own computer to the NEPS server system. The data is located within a so-called data enclave in which it is not possible to copy or store data on the researcher's own computer. The major difference to data access via the workrooms for scientific researchers in the NEPS building is that the data enclave does not allow us to control what users are doing in front of their desktops. Consequently, the data offered via remote access is less detailed. If the data offered via remote access is not sufficient, the user needs to work via remote execution. In remote execution, researchers send their syntax to the NEPS and get back results after checking them for their disclosure risk (see Chap. 20, this volume). The main handicap of remote execution is that researchers never have direct access to the original data. As a result, they can only improve their calculations afterwards by checking the received results. In general, the lowest level of technical data protection is realized in Scientific Use Files (SUF) offered via CD-ROM or download to the scientific community. In this case, after the data has left the NEPS coordination center, there is no longer any chance of controlling the data flow of the files. Scientific Use Files therefore contain less detailed data compared to data files offered via the other techniques of data access.

The NEPS has been set up to collect and disseminate educational data to the scientific community. In addition, it has to secure the data of all participants. The portfolio approach builds up a high-level multidimensional data protection system that still allows extensive data use for researchers.

19.7 Conclusion

In terms of data protection, the biggest challenge emerging from the complex multicohort sequence design of the NEPS is how to handle the collection, preparation, and dissemination of data appropriately. The procedures developed within the NEPS meet not only the requirements of the decisive data protection regulations, in particular, that of the respondents' privacy, but also the fundamental scientific need for high data utility.

Altogether, data protection ranks high within the NEPS. It is therefore a pivotal task for the coordination center that frames all the activities of the scientists and nonscientific staff working together within the NEPS.

Acknowledgments: For their tireless support, we wish to thank the University of Bamberg Data Protection Officer, the Bavarian State Commissioner for Data Security and his team, the Ministries of Education in the 16 federal states (involved in data protection issues in the school cohorts), and the data-collecting institutes, especially Heiko Sibberns (IEA Data Processing and Research Center, Hamburg) and Dr. Jacob Steinwede (infas, Bonn).

Endnotes

1 see http://www.projekt-datenschutz.de (Retrieved 7 Oct. 2010).

2 "The purpose of this Act is to protect individuals against infringement of their right to privacy as the result of the handling of their personal data."

3 "Every person shall have the right to free development of his personality insofar as he does not violate the rights of others or offend against the constitutional order or the moral law."

4 "Human dignity shall be inviolable. To respect and protect it shall be the duty of all state authority."

5 " 'Personal data' shall mean any information concerning the personal or material circumstances of an identified or identifiable natural person ('data subject')."

6 "Arts and sciences, research and teaching shall be free."

7 "racial and ethnic origins, political opinions, religious or philosophical beliefs, or trade-union membership … as well as data on health or sex life" [translated by the authors].

8 "as long as special circumstances do not require another form [of consent]" [translated by the authors].

9 "shall mean replacing the data subject's name and other identifying features with another identifier in order to make it impossible or extremely difficult to identify the data subject."

10 "the alteration of personal data so that information concerning personal or material circumstances cannot be attributed to an identified or identifiable natural person or that such attribution would require a disproportionate amount of time, expense, and effort."

11 "Persons employed in data processing shall not collect, process, or use personal data without authorization (confidentiality). … The obligation of confidentiality shall continue after their employment ends".

References

Bizer, J. (2007). Modernisierung des Datenschutzes: Vier Säulen des Datenschutzes. *Datenschutz und Datensicherheit, 31,* 264–266.

Brocks, H. (2009). *Praxishandbuch Schuldatenschutz* (2. überarbeitete Aufl.). Kiel: Unabhängiges Landeszentrum für Datenschutz.

Frick, J. R., Goebel, J., Haas, H., Krause, P., Sieber, I., & Engelmann, M. (2010). *Verfahren für den Datenschutz beim Zugang zu den SOEP-Daten innerhalb und außerhalb des DIW Berlin*. http://www.diw.de/documents/dokumentenarchiv/17/diw_01.c.347090.de/soep_datenschutzverfahren.pdf. Retrieved 7 Oct 2010.

Häder, M. (2009). *Der Datenschutz in den Sozialwissenschaften: Anmerkungen zur Praxis sozialwissenschaftlicher Erhebungen und Datenverarbeitung in Deutschland* (Working Papers No. 90). Berlin: Rat für Sozial- und Wirtschaftsdaten.

Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Lenz, R., Nylor, J., Schulte Nordholt, E., Seri, G., & De Wolf, P.-P. (2010). *Handbook on statistical disclosure control*. http://neon.vb.cbs.nl/casc/.%5CSDC_Handbook.pdf. Retrieved 10 Nov 2010.

Ichim, D., & Franconi, L. (2010). Strategies to achieve SDC harmonisation at european level: Multiple countries, multiple files, multiple surveys. In J. Domingo-Ferrer & E. Magkos (Eds.), *Privacy in statistical databases* (pp. 284–296). Berlin: Springer.

Institut für Demoskopie Allensbach. (Ed.). (2009). *Zu wenig Datenschutz? Die meisten sind mit persönlichen Daten vorsichtiger geworden* (Allensbacher Berichte Nr. 6). http://www.ifd-allensbach.de/pdf/prd_0906.pdf. Retrieved 7 Oct 2010.

Lane, J., Heus, P., & Mulcahy, T. (2008). Data access in a cyber world: Making use of cyberinfrastructure. *Transactions on Data Privacy, 1,* 2–16.

Metschke, R., & Wellbrock, R. (2000). *Datenschutz in Wissenschaft und Forschung*. Berlin: Verwaltungsdruckerei Berlin.

Rubin, D. B. (1993). Discussion statistical disclosure limitation. *Journal of Official Statistics, 9,* 461–468.

Shlomo, N., Tudor, C., & Groom, P. (2010). Data swapping for protecting census tables. In J. Domingo-Ferrer & E. Magkos (Eds.), *Privacy in statistical databases* (pp. 41–51). Berlin: Springer.

Relevant German laws and adjudications

Bayerisches Datenschutzgesetz (BayDSG) vom 23. Juli 1993, zuletzt geändert durch Gesetz vom 27. Juli 2009, Stand: Mai 2010.

Bundesdatenschutzgesetz (BDSG) vom 20. Dezember 1990, neugefasst durch Bekanntmachung vom 14. Januar 2003, zuletzt geändert durch Gesetz vom 29.07.2009, durch Artikel 5 des Gesetzes vom 29.07.2009 und durch Gesetz vom 14.08.2009, aktualisierte, nicht amtliche Fassung, Stand: 11. Juni 2010. English Version: Federal Data Protection Act (BDSG) in the version promulgated on 14 January 2003, last amended by Article 1 of the Act of 14 August 2009, in force from 1 September 2009.

BVerfGE, Entscheidungen des Bundesverfassungsgerichts, 65, 1. Urteil des Ersten Senats vom 15. Dezember 1983.

Grundgesetz für die Bundesrepublik Deutschland (GG), vom 23. Mai 1949, zuletzt geändert durch Gesetz vom 21. Juli 2010. English Version: Basic Law for the Federal Republic of Germany (GG) in the revised version, as last amended by the Act of 29 July 2009.