當代教育研究季刊　第二十三卷　第四期，2015 年 12 月，頁 009-040
DOI : 10.6151/CERQ.2015.2304.01

ESTABLISHING A COMPREHENSIVE LARGE-SCALE DATA INFRASTRUCTURE FOR EDUCATIONAL RESEARCH: THE EXAMPLE OF THE GERMAN NATIONAL EDUCATIONAL PANEL STUDY

Jutta von Maurice [1*] **Daniel Fuß** [2] **Hans-Günther Roßbach** [3]

ABSTRACT

Educational research profits greatly from the availability and analysis of longitudinal survey and competence test data. As the collection of such data and the preparation of a corresponding large-scale database is a challenging and expensive task, the results of these efforts ought to be shared among the scientific community. This article introduces the illuminating example of the German National Educational Panel Study (NEPS) and provides an insight into some of the key processes implemented in preparing and disseminating its rich empirical data to researchers from different disciplines and from all over the world. First, a short overview of the design and research topics of the NEPS is presented. The NEPS thus has a clear focus on competence development and educational processes—taking into account the relevant life-course-specific learning environments as well as issues of social inequality and educational decisions. Other major aspects include the special situation of persons with migration background and the various returns to education. As part of six different starting cohorts—ranging from early childhood to late adulthood—about 60,000 participants are interviewed and tested regularly

* Jutta von Maurice (corresponding author), Executive Director of Research, Leibniz Institute for Educational Trajectories.
E-mail: jutta.von-maurice@lifbi.de

Daniel Fuß, Operational Manager Research Data Center, Leibniz Institute for Educational Trajectories.
E-mail: daniel.fuss@lifbi.de

Hans-Günther Roßbach, Director, Leibniz Institute for Educational Trajectories.
E-mail: hans-guenther.rossbach@lifbi.de

Manuscript received: May 31, 2015; Modified: September 01, 2015; Accepted: September 16, 2015

Contemporary Educational Research Quarterly
December, 2015, Vol. 23, No. 4, pp. 009-040
DOI : 10.6151/CERQ.2015.2304.01

in different survey modes. Second, the article describes the most relevant work processes for the creation of Scientific Use Files—for example, data cleaning and editing, coding and variable generation, documentation and metadata management, as well as data enrichment. The data protection and dissemination strategies are also explained here. Third, the article finally also gives some basic information about data usage as well as an outlook on future developments within the NEPS.

Keywords: *german national educational panel study, research data infrastructure, educational research, longitudinal data, data dissemination, scientific use file*

The PISA Shock and the Demographic Situation in Germany

In Germany, the first results from the large-scale study PISA (Programme for International Student Assessment; OECD, 2001) caused a very intense reaction. The comparatively low scores of German 15-year-olds in all competence domains under study—mathematics, reading, and science—led to a rather emotional reaction labeled as the "PISA shock" in the media. Accordingly, a broad discussion broke out in politics and research and in the general public on the basis of these (unexpected) results. It is important to mention that Germany had not participated in any international large-scale assessments in the 1970s and 1980s and—because of this—was not prepared for these results. Since the millennium years, however, a lot has changed and German scores have been improving substantially (see also Table 1).

Table 1
PISA Results of Germany

	Mathematics	Reading	Science
Germany 2000	490	484	487
Germany 2006	504	495	516
Germany 2012	514	508	524

Note. Results retrieved from OECD (2007, 2014), Stanat et al. (2002).

The poor PISA 2000 results of German 15-year-olds must be viewed in light of the demographic structure (see Fig. 1).

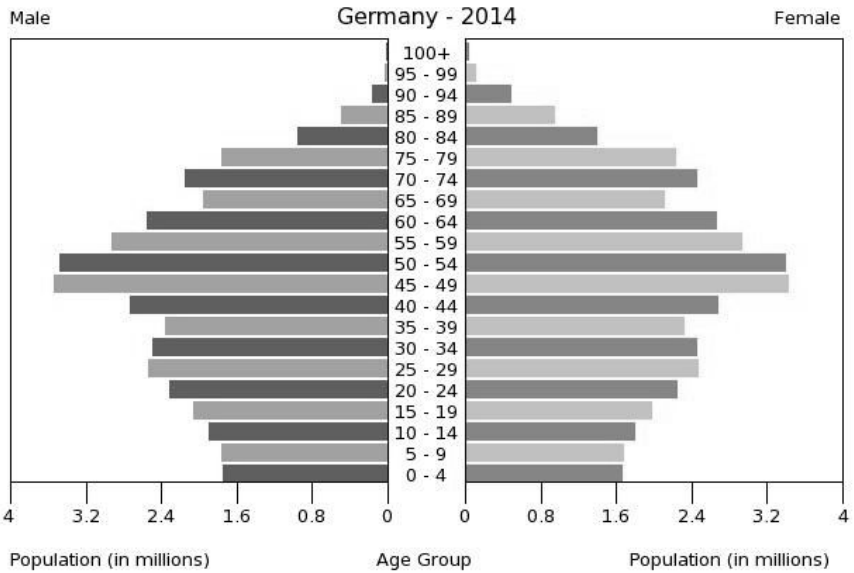

Figure 1. Demographic structure in Germany for 2014. Adapted from population pyramid by Central Intelligence Agency (2014).

Within Germany, we can see drastic demographic changes as a result of an increasing life expectancy and declining birth rates. Combined with decreasing unemployment rates (9.6% in 2000 and 6.9% in February 2015; Bundesagentur für Arbeit, 2012, 2015) as well as an increasing need for a well-trained workforce, these changes make it necessary to invest in education, as such human capital is indispensable for economic success. Germany—just like many other highly developed economies—has to invest in the education of younger people to introduce them to the labor market equipped with the best education possible, but also to take care of the education of their middle-aged and old-age population in order to keep them engaged in learning and to keep them up-to-date.

Let us return to the PISA shock. Research, politics, and all practitioners interested and involved in the field of education have learned a lot from these results—for example, about the strength of social disparities and the disadvantages of migrants in Germany. But what happens before the age of 15? At what age and where do these disadvantages evolve: within families, within schools, or somewhere totally different? The same discussion reappears after the age of 15: Are high competencies at age 15 a guarantee for later vocational

success and life satisfaction, or can low competencies or missing school degrees be caught up in later life phases? Or even—to argue quite provocatively—under what circumstances can a high income also be reached with medium to low competencies and with little or no formal qualification? But not only earlier and later status measures are needed; rather, processes and developments between different time points have to be plotted in great detail. A life-course perspective, a clear focus on the process of education and competence development, and a longitudinal design are hence our backbone in order to follow up individual trajectories and to analyze causal mechanisms (see Blossfeld & von Maurice, 2011).

Until now, much educational research has been focusing on learning in formal learning contexts. As Fig. 2 clearly shows for Germany, formal education hardly reaches people in their mid-twenties and older. Therefore, it became clear to us very early on that a comprehensive data set would have to go beyond this and cover education also during early childhood and adulthood.

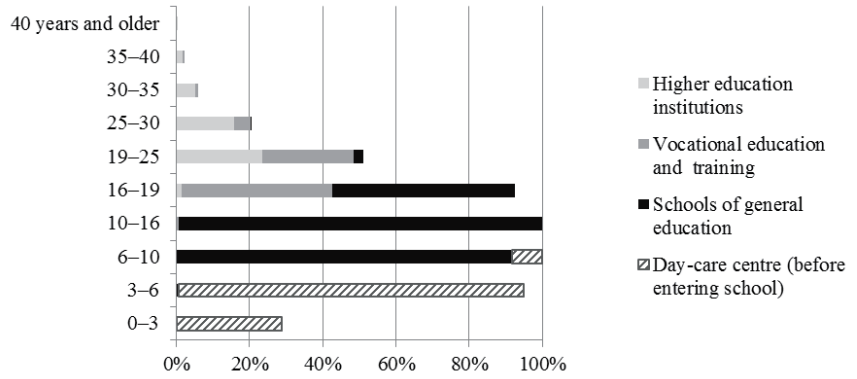

Figure 2. Participation in formal education in Germany in 2012/2013. Figure generated on the basis of data published as part of the National Report on Education 2014.

Longitudinal Data in Educational Research

The unexpected PISA results and the clear lack of longitudinal educational data in Germany in combination with the given constellation of declining birth rates, decreasing unemployment rates, and a growing need for well-trained employees formed a positive basis for initiatives in educational research, so that finally the idea of building up a broad database on

competence development and educational processes over the life span emerged in the early millennium years. Inspired by a clearly defined need on behalf of the German Ministry of Education and Research, a network of researchers was brought together to further elaborate on the idea of an integrated data set on competence development and educational processes covering the whole life span from early childhood until late adulthood. In this way, the idea of a cohesive National Educational Panel Study (NEPS) for Germany was born (for an overview see Blossfeld, Roßbach, & von Maurice, 2011). Various disciplines discussed and constantly concretized this idea.

The following sections give some information concerning the design and the major dimensions of the NEPS. Aspects of sampling, fieldwork, and data protection are also elaborated. As the NEPS is set up as an infrastructure facility for the social sciences issues of data preparation, the different steps of coding, variable generation, and documentation are also introduced in some detail. Finally, data dissemination strategies, data linkage activities, and user support offerings are described. It is shown that all these activities lead to research projects that are carried out by an impressive number of data users.

Analyzing the Life Span in a Multicohort Sequence Design

After the decision was made to adopt a longitudinal approach, a multidisciplinary team worked on the specification of the survey design. Following one large cohort of newborns over the life span is an amazing scientific challenge (carried out in an exemplary way by the Millennium Cohort Study, see for example; Dex & Joshi, 2004). In the discussion with possible funders of such a large longitudinal data set we soon realized that scientific results would have to be delivered quite rapidly and that politics were interested in evidence-based advice. Moreover, the expert teams involved in designing the plans also strived for instant data availability concerning their individual research focus. The selected multicohort sequence design brings together both aims: To follow educational processes and competence development from early childhood to late adulthood and to deliver results on pressing questions in education quite quickly—if you can call longitudinal results “quickly” at all. Our design includes six starting cohorts (SC), sampled and first surveyed at well-selected points in life—especially before or immediately after central transitions between important formal learning environments:

- SC1: Sample of 7-month-olds,
- SC2: Sample of children in Kindergarten 2 years before regular school enrolment,
- SC3: Sample of fifth graders,
- SC4: Sample of ninth graders,
- SC5: Sample of first-year students in universities or universities of applied sciences,
- SC6: Sample of adults (aged 23 to 64) independent of their educational and vocational situation.

The six starting cohorts are illustrated in Fig. 3.

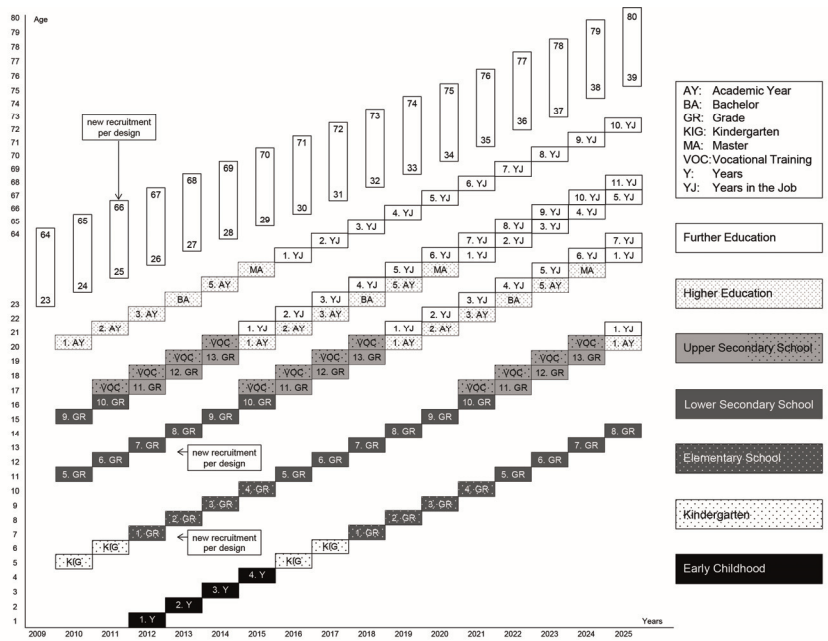

Figure 3. Multicohort sequence design of the National Educational Panel Study.

The implementation of this design was funded for a 5-year period (2009-2013) by the German Ministry of Education and Research; after a successful evaluation by the German Council of Science and Humanities the

NEPS finally received permanent funding within the Leibniz Institute for Educational Trajectories as a research-based infrastructure facility under the umbrella of the Leibniz Association.

Major Dimensions as a Guideline for Instrument Development

In order to provide data in different cohorts that can be related to each other for educational analyses across the life span, we decided very early on to focus primarily on five dimensions:

- Competence development across the life course: We especially follow the development of domain-specific competencies across the life span (Artelt, Weinert, & Carstensen, 2013; Weinert et al., 2011) because competencies are not only one of the major determinants of educational, vocational, and also private life paths but are also a benefit per se of any type of educational process. Thus, we focus on reading, mathematics, and the natural sciences. It is a challenging task to link the competence tests developed over the life span and—in this way—make the development of domain-specific competencies visible over the life span. The domain-specific competencies are supplemented with the assessment of domain-general cognitive functioning, meta-competencies such as metacognition and ICT literacy, so-called L1-tests (tests in the participants' first language administered to selected migrant groups), and finally—at least in some of our starting cohorts—stage-specific competence tests (e.g., precursor competencies in the early childhood cohort, orthography in elementary school, and English in secondary school). We also started to test vocational competencies in participants in vocational training or higher education. However, because of the large number of different vocations to be considered in a panel such as NEPS, this has been a very challenging task. Besides cognitive variables, some basic "non-cogs" are included in our design, such as interests and motivation, personality, and selected social behavior indicators (Wohlkinger, Ditton, von Maurice, Haugwitz, & Blossfeld, 2011).
- Education processes in life-course-specific learning environments: Learning and competence development cannot be understood without a thorough look at the various learning environments that people live and engage in. Therefore, formal (e.g., schools), nonformal (e.g., corporate trainings), and

informal (e.g., learning in peer groups) learning environments have to be considered (Bäumer, Preis, Roßbach, Stecher, & Klieme, 2011). The family—as one of the earliest and environments that also lasts the longest—is given special attention. In the operationalization of this dimension, aspects of quantity and quality are included. Using the data collection of the NEPS, researchers are able to analyze additive as well as interactive effects of subsequent and parallel environments.

- Social inequality and educational decisions across the life course: The PISA results have shown that great social disparities still persist among 15-year-olds in Germany. This has been discussed especially by politics and by the general public. Within NEPS, we collect very detailed information on how primary and secondary effects of familial social background may affect educational trajectories. Measures of social and cultural capital are included all over the life span (Stocké, Blossfeld, Hoenig, & Sixt, 2011).
- Education acquisition of persons with migration background: The PISA results also showed us quite clearly that migrants in Germany experience severe disadvantages. We wanted to learn more about whether—and if so, why—primary and secondary effects evolve (or fade) over the life span. Therefore, we do not concentrate on the first generation of migrants; instead, we provide a detailed migration biography of the family so that we can trace back our targets' background to the generation of their grandparents. In this way, we can see how migration effects persist over generations. The above-mentioned tests in the first language (L1) and tests of cultural knowledge in selected migrant groups can enrich our analyses and give a closer look at possible mechanisms (Kristen et al., 2011).
- Returns to education across the life course: Education does not only lead to competence development (at least if it is effective) but might also have other benefits beyond an increasing competence level. Within the "hard" economic returns to education, data on income, wealth, acquisition of certificates, and status are central. These indicators are supplemented with data that allow us to analyze also the effects of education on health, satisfaction, social and political participation—which are often labeled as "soft" noneconomic returns (Gross, Jobst, Jungbauer-Gans, & Schwarze, 2011).

While we follow all those dimensions from early childhood through to late adulthood, the respective operationalization in each of our studies

conducted has to adapt to the age and educational stage under study: The family environment of newborns, for example, can be characterized by parental activities such as looking at picture books with the child, in elementary school homework support might be effective, and when leaving school support in finding an apprenticeship position arises as a new topic. In adulthood, family formation emerges as a totally new learning context. Moreover, all above-mentioned dimensions must be supplemented with stage-specific questions: How do target persons master the transition from Kindergarten to elementary school (and are schools effective in supporting parents in their decisions)? At what point are vocational decisions taken by adolescents (and what role do their parents and networks play in this process)? Why do some adults engage in lifelong learning and others do not (and how can employers positively influence the decision to take up further training)?

In the context of instrument development, it is especially important to work in a highly interdisciplinary way. It is only through intense collaboration that we can provide data that will allow researchers from very different disciplines to answer their research questions.

Sampling, Surveying, and Tracking Study Participants

We started our panel cohorts in 2009 (SC6), 2010 (SC2, SC3, SC4, SC5), and 2012 (SC1). All cohorts are representative random samples. Whenever possible, we drew our samples from institutions because surveying and testing children, adolescents, or adults in Kindergarten, school, or university contexts is cost-effective on the one hand and leads to data that may contribute to answering central questions on learning contexts (e.g., on effects of teacher characteristics or class composition or competence development). Overall, more than 60,000 participants were included in NEPS in the first waves of the six cohorts. Additionally, parents and educational staff were included in the younger cohorts (see Table 2).

Table 2
Starting Cohorts of the National Educational Panel Study

Starting cohort	Targets	Sampling	N (Targets)	Context persons
1	Early childhood	Individuals	3,439	Mothers, childminders
2	Kindergarten	Institution-based	3,007	Parents, educators, principals
3	Grade 5	Institution-based	6,112	Parents, teachers, principals
4	Grade 9	Institution-based	16,425	Parents, teachers, principals
5	First-year students in higher education	Institution-based	17,911	–
6	Adults	Individuals	13,576	–

The survey modes and survey designs for the target respondents are adapted to the different ages. In this paper only limited information can be given on this topic. For more information please refer to the data and documentation section of our homepage.[1] Before school entrance, only one-to-one test situations with the targets are feasible. The tests carried out with very young children (starting with 7-month-olds) mostly rely on video observation recording their behavior in semistandardized situations. Tablet-based testing is administered to children starting at the age of 3. In Kindergartens, the testing situation is still one-to-one but designed in a more "play-like" fashion that mostly avoids active language production by the preschoolers. In schools, we use group tests (starting with smaller groups and adapting group size as the children are getting older) and apply more traditional testing methods (until now, paper-based). In Grade 3 we also start handing out questionnaires to our target students. After our targets leave the institutional context we conduct CATI interviewing (computer-assisted telephone interview) and individual testing in CAPI (computer-assisted personal interview) or online mode. Also, adults are included in a mixed CATI and CAPI mode. Besides the age of the person under study—and the fit between mode and individual capacities, expectations, and life situation—

[1] https://www.neps-data.de/tabid/294/language/en-US

cost-arguments and comparability of data between different waves within one cohort or between cohorts must always be balanced out. Test and interview lengths vary considerably between the ages, from half an hour or less (prior to school enrolment) up to 4 hours (in higher school classes). Especially when tests and surveys are carried out outside the institutional context, our survey time is not only limited by the capacity of the target persons but also by the time that they are willing to invest in study participation. For example, in the adults cohort we try to avoid survey times exceeding 1 hour to 1.5 hours in order to keep panel stability high. In most cohorts testing time for one competence domain is 30 min; the survey part varies, in most cases, between 20 and 60 min.

For context persons the survey mode also varies: Parents are interviewed primarily by CATI; it is only in the early childhood cohort that we use CAPI interviews (as we are in the home environment anyway). Kindergarten and school staff are always included in the PAPI (paper-and-pencil interview) instruments.

All data collection is conducted by two highly experienced survey institutes.

After sampling the participants (and context persons) and Wave 1 data collection, we invested a lot in tracking and panel care. Keeping addresses up-to-date through active address inquiry or by offering an online portal for address updates, telephone hotlines at the data-collection institutes and at our scientific institute, regular flyers and brochures about our work, giving out carefully selected incentives to the participants, regular greeting cards from the director of our institute, and so on, are just some examples of our concept for fostering panel stability. It is important to note that we are especially interested in those children, adolescents, and adults that do not follow the traditional paths. Therefore, we track and survey all participants independent of their educational development. Children who, for example, leave their school context due to grade repetition, grade skipping, or simply as a result of their family moving away are followed individually (but with a reduced testing program).

Data as a Valuable Resource for Educational Research

The development of the longitudinal design was linked to a quite consensual commitment that the collected data should be made available as a

common good to all researchers interested in educational research. There is an intense discussion in Germany and in other countries concerning data dissemination. In the following, we are going to mention a few German milestones only.

- The value of data: One of the milestones in Germany was the clear commitment to the view that research data are valuable and worthy of long-term storage. The German Research Foundation appointed an international commission on professional self-regulation in science already in the late 1990s. This commission agreed on several recommendations concerning "Safeguarding Good Scientific Practice". In this memorandum, scientists placed a special emphasis on primary data, specifying that "[p]rimary data as the basis for publications shall be securely stored for ten years in a durable form in the institution of their origin" (Deutsche Forschungsgemeinschaft, 2013, p. 74). It is only through the long-term storage of primary data that analyses and results can be replicated.
- Open access to data: One major step behind the pure storage of data was the commitment to enabling free access to data. In the "Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities" (2003) this is stated very clearly:

  > Our mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society […] Establishing open access as a worthwhile procedure ideally requires the active commitment of each and every individual producer of scientific knowledge and holder of cultural heritage. Open access contributions include original scientific research results, raw data and metadata, source materials, digital representations of pictorial and graphical materials and scholarly multimedia material. (p. 1)

- Need for research data infrastructures: In 2004, the German Data Forum (RatSWD) was established by the German Federal Ministry of Education and Research. According to the homepage, their main goal is "to sustainably improve the research data infrastructure underlying empirical research and contribute to its competitive ability on an international level" (RatSWD, 2015). In the work carried out by the German Data Forum the secure storage and easy access of data are of outstanding importance.
- Building up research data infrastructures: On the basis of this fundamental value given to data and the commitment to free access, the German Council of Science and Humanities (Wissenschaftsrat) recommended in 2011:

> Therefore the Council recommends to infrastructure operating research institution the comprehensive and long-term archiving of quality-tested data which are relevant for the respective scientific communities. Regarding quantitative primary research data, the German Council of Science and Humanities recommends enhancing the user friendliness of the services in archiving and researching data even more intensively. (p. 56)

Finally, they stress the importance of research infrastructure in a clear statement:

> Successful operators of a research infrastructure that inspire new research topics and whose data are used to achieve research results that win international recognition should not, in terms of reputation, be left behind the users of this infrastructure who are successful in research with these data. (p. 82)

- Commitment at ministerial level: Within this line of argumentation the Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung, BMBF) states on its homepage (as of March 8, 2015):

  > Researchers need modern, strong research infrastructures to produce efficient, internationally competitive research. The BMBF has launched specific funding measures for the humanities and social sciences in order to encourage and promote the development and expansion of research infrastructures across national borders also in these areas.

The first argument often mentioned in relation to data dissemination for the scientific community—especially in large-scale and longitudinal data sets—is the high cost of data collection. But there are more aspects to be discussed: Designing and conducting surveys such as the one described earlier requires expertise from different disciplines. It is not easy to bring together such interdisciplinary networks and to motivate them to engage in the production of a big data set (including all the organizational and operational work that comes along with such an endeavor). This “pool of expertise” must be “exploited” by as many researchers as possible in order to achieve the greatest possible scientific progress. Finally, the investment of all survey participants must be considered: It is a time-consuming task for the participating institutions to help us organize our data collection. What is more,

we cannot appreciate too highly that people open up their homes to give us interviews and even participate in our demanding competence tests (only to accept more often than not the frustrating experiences when realizing that they cannot solve a large amount of the items presented to them). So it is the combination of arguments related to cost, striving for scientific progress, and the responsibility for our participants that speak for data dissemination. Therefore, broad data dissemination was defined as a fundamental funding condition right from the start of the NEPS.

Data are thus disseminated as so-called Scientific Use Files (SUF). These files make available the entire database to any researcher who is interested in working with these data for the purpose of scientific research and who meets the formal criteria as described in the following section. In the case of NEPS, for each starting cohort an individual Scientific Use File is available including edited and documented information from the most current and all previous panel waves. Access to all NEPS Scientific Use Files is granted free of charge.

Data Protection

A thorough preparation and editing process is inevitable to make available the information collected by the NEPS survey in the form of high-quality panel data ready for scientific use. A set of complex steps including data anonymization, cleaning and editing, coding and variable generation, documentation and registration, as well as data enrichment has to be conducted before any Scientific Use File can be released to the research community. This requires great efforts by the responsible Research Data Center at LIfBi (RDC LIfBi), but also by other operational units contributing to the NEPS. Whereas the design of user-friendly data structures ensuring a maximum of data usability represents one central concern, another objective is to guarantee a maximum of confidentiality protection for the survey respondents and their individual microdata. These partly contrary guiding principles are reconciled by an elaborate portfolio approach that combines five modules of data security measures. It follows the work of Julia Lane, Pascal Heus, and Tim Mulcahy (2008; see also Meixner, Schiller, von Maurice, & Engelhardt-Wölfler, 2011) and allows us to not only supply the highest level of information extent possible but also to protect user anonymity according to national and international standards.

- Organizational measures: The NEPS research data infrastructure has been established for use by the scientific community only. Data users must first prove their identity as scholars, researchers, or academic staff. This limitation to persons with an academic affiliation makes unauthorized purposes of data use improbable and also increases the trust and acceptance on the part of study respondents, as their information is handed over exclusively to persons of integrity.
- Legal measures: Researchers and data-providing agencies in Germany are bound by numerous legal regulations concerning data protection. The NEPS data use agreement specifies not only the legal rules data users have to comply with but also the scientific purpose of the intended research project. Signing the agreement is a necessary prerequisite for any NEPS data access.
- Educational measures: Another module refers to the establishment of a community of well-trained and trusted users. Specific documentation materials and regular training courses are crucial elements to sensitize users for issues of data security and to make them fully aware of their responsibilities when working with sensitive microdata. For particular forms of data access it is mandatory to personally attend a NEPS training course.
- Technical measures: Technical provisions for data protection are primarily related to different modes of data access. NEPS offers three alternatives that vary both in terms of control over data usage and in the level of anonymization. That means, more sensitive data are available under secured conditions only, and any use of these data is subject to strict controls by the NEPS staff.
- Statistical measures: All disseminated NEPS data are de facto anonymous data. This implies coarsening or cutting off identifiable information to minimize the risk of statistical disclosure. String variables relating to openly asked questions are checked carefully. Further applied techniques of statistical data protection include changes to the data itself, ranging from compressing or aggregating information (e.g., by not disclosing somebody's occupation but only the branch of the economy) to adding noise (i.e., true values are changed following a given model). However, the general objective is to make available our collected data containing as much explanatory power as possible.

Data Preparation

The preparation of huge collections of empirical information embedded in a complex panel design such as the NEPS survey requires systematic and collaborative efforts of data cleaning and data editing. Checking, correcting, and integrating data material to create Scientific Use Files is a major task of the RDC LIfBi. It also coordinates contributions from the field institutes and from other operational units participating in NEPS, who are also involved in these processes as they are highly familiar with the data collection as well as with the instruments and the theoretical constructs implemented.

- Data cleaning: Cleaning the data, in particular, is realized as an iterative process whereby data errors or inconsistencies are conjointly identified, negotiated, and corrected. In fact, two preliminary data versions are usually disseminated within the NEPS network for control, feedback, and approval before the final Scientific Use File is released to the scientific community. This strategy facilitates a careful examination of the data by many researchers and at different stages of the editing process.
- Data editing: The editing process also takes place in a multieditor environment that has been set up to commit all coworkers to some basic principles. Besides leaving raw data unchanged and organizing the data editing process into intermediate steps, another fundamental principle is to make sure that the whole process is replicable. This is achieved by completely syntax-based procedures using the software Stata as a standard technology, supplemented by a version control program recording who has changed what in the respective syntax elements. As a result, the preparation of a Scientific Use File remains traceable at any point in time as it is being documented in high precision. To give an example: For the first Scientific Use File of SC6 (adults, doi: 10.5157/NEPS:SC6:1.0.0), up to nine persons—predominantly from the Research Data Center—were involved in the data preparation process; together they produced more than 150,000 lines of syntax code and documented almost 900 code revisions.
- Data integration: Due to annual or semiannual data collection sweeps along the six different NEPS starting cohorts, the handling of panel-wave data (just like data merging or harmonization of variables) would soon turn into a serious and exhausting task for researchers. Thus, in order to alleviate the users' data-management efforts, integrated longitudinal data sets (usually in

long format) are provided for all cohorts. For instance, rich life-course data collected retrospectively and prospectively over several waves are being integrated and harmonized into easy-to-analyze episodes/spell data sets. This considerably reduces the burden of data management on the part of the researchers. The same applies to the provision of homogeneous data and label structures across waves and cohorts.

- Data registration: The final step of the data-preparation process consists of the assignment of a unique and persistent digital object identifier—a so-called DOI—to each Scientific Use File (Wenzig, 2012). DOI assignment is carried out at da|ra, the German registration agency for social and economic data.[2] By using the DOI researchers are able to cite NEPS data directly in their publications in a very easy and precise way. This will ensure the traceability of the research process; a demand that has become more and more important in the context of good scientific practice. The DOI not only indicates the relevant NEPS starting cohort and the data version of the Scientific Use File, it also refers to a landing page at the NEPS web portal that provides basic metadata relating to the data and describing ways of data access. Users can find the DOI—together with a short, human-readable comment—written as a label for each data set of a Scientific Use File.

Coding and Variable Generation

A routine task within the data editing process is the recoding of textual responses to given categories as well as the coding of open answers. The NEPS survey includes a number of pieces of such information in an open string format. Whenever respondents are invited to state basically anything they want, a practicable way to deal with this kind of entry is coding—that is, the assignment of a numerical code from a selected category scheme or classification to the string information. In the context of NEPS, the most complex and elaborated string coding procedure pertains to occupations. Although some automatic procedures have been developed (e.g., string comparison with keyword lists, suggestion-based pre-coding of strings), coding is mostly done manually by experts with a significant amount of occupational knowledge and methodological know-how. In the end, all NEPS

[2] http://www.da-ra.de/en/home/

data are provided with classical codes for occupations and vocational education, such as the German classification of occupations (KldB-88 and KldB-2010, see Paulus, Schweitzer, & Wiemer, 2010) and the International Standard Classification of Occupations (ISCO-88 and ISCO-08, see ILO, 2013). Business sectors and industries, adult and further education courses, as well as fields of study in higher education are also coded according to standard schemes. To illustrate the amount of work that this entails: Over the course of preparing the current Scientific Use File of SC6 (adults, doi: 10.5157/NEPS:SC6:5.0.0), a package of more than 100,000 occupation-related codes had to be assigned to open entries and counter-checked for coding quality. Moreover, information related to locations such as place of residence, place of birth, place of employment, or place of vocational training is converted into categorical variables by using official district numbers.

The generation of additional variables further improves the quality and utility of NEPS data. By default, Scientific Use Files are delivered with scales and generated variables that measure socioeconomic status and occupational prestige—for example, the Magnitude-Prestige scale (Wegener, 1985), the SIOPS-88 and SIOPS-08/Treiman scale (Treiman, 1977), the International Socioeconomic Index of Occupational Status (ISEI-88 and ISEI-08; Ganzeboom, de Graaf, & Treiman 1992; Ganzeboom, 2010), the EGP classes (Erikson, Goldthorpe, & Protocarero, 1979), and the occupational class scheme by Blossfeld (1985). To indicate the level of educational attainment, variables for CASMIN and ISCED-97 are provided. With regard to spatial information, several variables containing codes for administrative regions (“Regierungsbezirke”) and Federal States (“Bundesländer”) according to the NUTS hierarchy of regional clusters are available. Last but not least, several other variables are generated addressing more specific issues such as personality traits or migration background (Olczyk, Will, & Kristen, 2014).

Documentation

Good scientific practice and research strongly depend on comprehensive and accessible data documentation. NEPS has established an innovative strategy of producing documentation materials in both German and English. It relies on a structured and integrated approach to metadata management (Wenzig, 2013). The majority of the numerous NEPS surveys include instruments (i.e., questionnaires or competence tests) that define dozens of

questions and items as well as filtering or interviewer instructions. Many of these items are repeatedly deployed across waves, but also across different cohorts. In consequence, an extraordinary abundance of metadata has to be administered in a way that efficiently links, deduplicates, reuses, and presents all relevant information. For that purpose, a relational SQL database has been developed. It enables storing and linking diverse metadata on studies, instruments, items, data sets, variables, and answer schemes in a systematic fashion. For example, questionnaire items are linked to data sets—thus allowing for a dynamic documentation that directly leads from the data set variable to the corresponding question within a questionnaire. The reuse of metadata provides for a tracking of inadvertent changes in variables across panel waves and starting cohorts. The central maintenance of NEPS metadata in one big database ensures a high documentation utility because every correction and extension becomes effective in a synchronous and consistent way in all derived documentation materials, such as codebooks, survey instruments, data set labels, or the interactive exploration tool NEPSplorer. The NEPSplorer is a software tool developed in-house that offers a full text search through all recorded metadata of survey instruments and Scientific Use Files. For each item, information on question phrases, corresponding variables, answer categories, authors, interview instructions, concepts, keywords, etc. can be browsed and—according to individual requirements—stored in a user's personal watch list. In addition to the innovative NEPSplorer, a number of standard written documentation files are available for download. As mentioned before, some of these documents are directly derived from the meta-database, such as survey instruments and codebooks. Other documents—such as data manuals, release notes, descriptions of competence tests, methods reports, and technical reports (e.g., on weights, anonymization, regional indicators)—are compiled and provided in order to reduce usage hurdles and to inform researchers about the strategies of how the data have been prepared and how to make use of them. Finally, the RDC LIfBi provides semantic data-structure files for an easy and intuitive data exploration without having to conclude a user contract beforehand. Semantic data-structure files are emptied Scientific Use Files containing all variables and labels, but no data rows.

Dissemination Strategies

In November 2012, the RDC LIfBi became an accredited research data center and member of the Standing Committee Research Data Infrastructure.

As such, it is obliged to act according to the criteria of the German Data Forum for the establishment of research data infrastructures (RatSWD, 2010). These criteria emphasize the user-friendly preparation and transparent provision of high-quality research data for scientific purposes, complemented by an extensive user service (e.g., website, individualized advice, data workshops and presentations). The above-described data editing processes as well as the subsequently outlined data dissemination strategies follow the basic principles as defined by the German Data Forum. Thus, the RDC LIfBi and NEPS, respectively, represent an integral part of the German research data infrastructure.

To make NEPS data available to the scientific community in a convenient way, a threefold model for flexible data access has been implemented. Research data are provided (a) by secure download from the NEPS website, (b) by an innovative remote-access technology called RemoteNEPS (Barkow et al., 2011), and (c) at on-site workstations at the RDC LIfBi in Bamberg. The different modes of data access support the full range of users' interests with regard to data utility while complying with strict standards of data confidentiality protection. All three alternatives grant access to Scientific Use Files with a common data structure, but they vary with regard to the degree of data anonymization. Data files that are available "on-site" provide more sensitive information than files that are available via "remote access"—which, in turn, contain more information than the "download" version of data files. In other words: The data provided for download from the website feature the highest level of anonymization, whereas data made available through the controlled environment of RemoteNEPS feature a moderate level of anonymization. On-site data feature the lowest level of anonymization. In the case of our school cohorts, this hierarchical concept means, for example, that information provided by the principals about their respective schools is available in RemoteNEPS and on-site only, whereas more fine-grained regional indicators are accessible only on-site at LIfBi in Bamberg.

Irrespective of the preferred data access mode, the use of NEPS data requires a signed data use agreement. As noted above, only persons who are members of a scientific institution are eligible to apply for data usage. By signing the agreement all users commit themselves to strict data protection guidelines that forbid any attempt of reidentification, passing on any data without permission, or using the data for other purposes than the specified research objective. In case of violated stipulations serious penalties are

prescribed in the contract regulations (e.g., high monetary penalty, proscription, exclusion from further data usage).

After the data use agreement has been approved, the researcher receives individual NEPS login details enabling him or her to download all available Scientific Use Files. The data access option of RemoteNEPS requires two further steps: signing a supplemental agreement and individual authentication through the researcher's keystroke biometrics. The same applies to on-site data access, which requires data users to stay as a guest researcher at LIfBi in Bamberg. Those scientists who want to analyze the highly sensitive on-site data are constrained to work within a controlled physical environment—that is, all input and output devices are locked down and the computers are not connected to the Internet or any other local area networks. RDC LIfBi staff are allowed to monitor any work being carried out with on-site data at all times. In order to prevent any copying or removing of sensitive data, each output from on-site or remote data usage passes through a review before being provided to the researcher.

Enrichment by Data Linkage

Linkage of data is an important source for increasing the range of research questions and the analytical potential of the data. Bringing together the rich longitudinal survey and test data from the NEPS study with administrative microdata or with regional macrodata from external sources significantly enhances the capability of empirical research. At the same time, data linkage is a multifaceted venture in methodological, technical, and data protection terms. Compared to many other countries there are very restricted options—based on arguments of data protection and the lack of a uniform personal identifier—for linking the NEPS data with official databases such as insurance data, health data, or school records. At this juncture, conceptual frameworks have been established at RDC LIfBi for two data linkage scenarios:

- Linkage with administrative data: NEPS data in selected starting cohorts are linked with administrative data on the basis of the respondents' explicit consent. The linkage takes place at individual level by means of probability matching relying on parts of the name, address data, and basic sociodemographic traits. Administrative data are provided by the German

Federal Employment Agency, comprising information on the complete employment and unemployment history (except self-employment), benefit recipient history, yearly gross income, participation in active labor policy measures, etc. Currently, enriched data of all published waves of SC6 (adults) are available for scientific analyses at special workstations in the Institute for Employment Research in Nuremberg. It is intended to further extend this service by successively linking survey data from other starting cohorts to administrative data.

- Linkage with regional indicators: On the one hand, NEPS data include fine-grained regional information up to street-section level that can be merged easily with the survey and test data. These macro-level data are purchased from two leading commercial providers of geodata in Germany (Microm, Nexiga). These two databases comprise a variety of indicators—such as type of residential neighborhood, residential density, share of cultural and leisure facilities, average purchasing power and status level, unemployment rate, family and age structure, confessional structure, proportion of foreigners and ethnic composition, mobility of neighborhood residents, and the dominant social milieu of the given neighborhood. On the other hand, the RDC LIfBi offers researchers the opportunity to merge their own regional indicators (e.g., from the German Federal Statistical Office) on the basis of municipality keys with NEPS data. For reasons of data protection these regional indicators are subject to a careful check before enrichment, and all analyses with linked regional data are restricted to secure on-site or remote data usage.

User Support Activities

As already pointed out, the development of a large-scale research data infrastructure does not end with the preparation, enrichment, and dissemination of the data. Another key task is to offer comprehensive services to the data users. The aim of facilitating work with NEPS data poses a significant challenge, as the longitudinal and hierarchical design of the survey creates a complex database with numerous data files. Reconciling this complexity with convenient and proper data usage requires a multitude of support activities.

At its core, training courses in German and English are held on a regular basis—about eight per year—at the LIfBi in Bamberg. These courses cover theoretical, methodological, and technical topics. These 2-day courses are divided into two sessions that can be attended either separately or together. Whereas the first day is reserved for a general introduction to the design of the study, the structure of data files, ways of accessing the data, and the use of documentation, the second day has a particular focus on a certain starting cohort and/or on selected methodological challenges—for example, the use of weights and competence test scores. An overall objective of both basic and advanced training courses is the sensitization of researchers to the terms and conditions of NEPS data usage, especially to issues of privacy and data protection. For that reason, course participation is obligatory for data users who want to enroll in the biometric authentication system for gaining access to RemoteNEPS and on-site data. Additional off-site courses represent an important building block in the strategy to further promote NEPS data to an international audience of researchers. Off-site courses are organized either in the context of national and international scientific events (e.g., summer schools, conferences) or as stand-alone events. For instance, NEPS user trainings were organized at the Educational Research Institute in Warsaw (Poland), the European University Institute in Florence (Italy), and at the National Taiwan Normal University in Taipei (Taiwan). In total, NEPS has welcomed far more than 500 researchers and students from Germany and abroad to its user trainings to date.

Further support activities by the RDC LIfBi include the maintenance of a comprehensive and bilingual website (www.neps-data.de) and the administration of an e-mail hotline (fdz@lifbi.de) as well as a phone hotline for individualized advice. The e-mail hotline—operated via electronic ticket system—plays a particular role in the interaction between data users and the RDC LIfBi. Besides dealing with questions concerning data handling and data access issues it serves to process all import and export requests from researchers using RemoteNEPS or working with on-site data. In 2014, almost 2,000 e-mail queries were answered, with almost 400 of those e-mails asking for imports or exports from the controlled data environment.

Data Usage

The continued release and dissemination of NEPS data over the last couple of years has been accompanied by a steady rise in data usage,

indicating the growing demand for longitudinal data on educational processes from early childhood to late adulthood in the national, but also international scientific community. Figure 4 depicts the development of (a) the number of research projects based on NEPS data and (b) the number of unique data users who are involved in at least one of these research projects. The graphs start in the summer of 2011 when the first Scientific Use File was released for the adults' starting cohort (SC6).

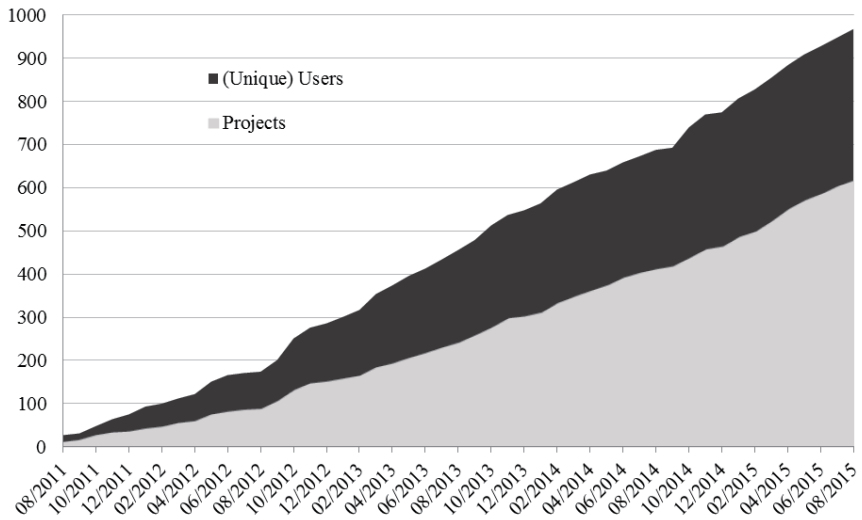

Figure 4. Development of NEPS data usage from August 2011 to August 2015.

By the end of August 2015, more than 600 research projects with almost 1,000 researchers involved had been registered. Currently, about 15 new project proposals are submitted to us on average per month, with a tendency to rise. Because of the growing recognition among the scientific community and the increasing attractiveness of NEPS over the course of forthcoming data releases it is to be expected that this trend will persist in the future. A remarkable aspect in this context is the relatively high share of research with NEPS data being carried out at institutions from abroad. Over 10% of all NEPS users work in other countries, that is, in Italy, the United Kingdom, the US, Finland, Switzerland, Austria, Australia, South Korea, India, etc. This reflects the high level of international usability of NEPS data, mainly ensured by our English documentation, but also the rich potential of this "national" survey data for international educational research.

A closer look at NEPS data usage shows us a broad spectrum of research questions and topics that are being addressed by scientists from various disciplines (e.g., sociology, psychology, educational science, political sciences, economics, demographics). This spectrum ranges from learning processes and competence development to international comparisons. A significant proportion of projects are dealing with issues of education and competencies in relation to labor market research, migration research, social mobility research, or family research. Other projects based on NEPS data address research questions from the fields of pedagogical psychology and educational sciences. Further projects are dedicated to methodological topics or to educational reporting and the provision of expert advice for policy makers. All released NEPS starting cohorts are subject to research projects, many of them conducted as part of academic qualification efforts such as doctoral projects.[3]

Politics and educational administration in Germany are strongly interested in results based on the NEPS data. Although they cannot use the data per se (data access is limited to researchers) they benefit greatly from the manifold output of scientific analyses. Information on findings and new insights are communicated to them through presentations, regular newsletters and—in the near future—by providing summaries of scientific articles in an edited form using plain language.

Outlook

The primary aim of the Leibniz Institute for Educational Trajectories and the NEPS Network is to follow the participants of all six panel cohorts along their individual life trajectories. Through regular surveys and tests we are gaining a constantly increasing insight into the educational processes and competence development of our study participants. Our current experiences show that changing life situations (e.g., entering or leaving the school context) are particularly critical points for panel stability. Despite our intense efforts in relation to panel care a certain amount of dropout cannot be avoided. Panel attrition rates and, especially, any signs of bias in the data caused by panel dropout have to be monitored very carefully. In the long run, we will have to decide on whether to adjust the data-collection design (e.g., only short

[3] For a complete list of all registered NEPS research projects including a brief description and the names of all involved persons see: https://www.neps-data.de/en-us/datacenter/researchprojects.aspx.

follow-ups every 2 or 3 years instead of intense surveying and testing on an annual basis) or even give up an entire cohort. On the other hand, the multicohort sequence design is open to a restart of cohorts. A restart is especially valuable when important changes in society or in the political background have occurred (e.g., a massive expansion of early child care services for children below the age of 3 in Germany initiated in 2008).

Another aim of the NEPS is to constantly improve our survey and test instruments. An increasing use of computer-based testing modes with branched or adaptive testing, online surveys, and the development of instruments for measuring new constructs (e.g., IT-learning or cultural knowledge) are scientifically challenging. Moreover, these innovations constantly help to improve the survey and test situation for the participants (e.g., by avoiding hard and thus sometimes frustrating test items) and to keep the panel interesting for them (e.g., by presenting participants with items adapted to their individual life situation), which is of utmost importance.

After having been in the field for five years, we are now seeing first initiatives to use the NEPS data as a benchmark. Researchers can survey and test highly selective groups (e.g., students with certain disabilities) or groups affected by some kind of "exposure" (e.g., school reform in a given Federal State within Germany) and then refer to the NEPS data as a representative comparison group. Another significant component of NEPS-based research projects are international comparisons using our data as a reference for the case of Germany.

With regard to our efforts in preparing and publishing NEPS data for scientific use by national and international researchers the focus is on continuously releasing six or more Scientific Use Files per annum—at least one for each starting cohort. These releases will include all available data so far, supplemented by new survey and test data from the most current wave, respectively. Thus, the potential of the data is going to increase steadily with respect to both addressing new research questions and employing more elaborated techniques of analysis. It is our primary intention to maintain our high standards of data editing, dissemination, documentation, and user support as described in this article. Of course, these processes are subject to permanent improvement in terms of user-friendliness and data usability. For instance, we aim to (a) reduce the time span between fieldwork and data availability, (b) further broaden the data portfolio by generating new variables and scales or by linking the survey data to other data sources, (c) expand data documentation

through new materials and further developed tools, and (d) intensify the contact with the research community through several measures, such as the continuation of our semiannual data newsletter, the establishment of a user forum on our website, the organization of a NEPS user conference, the provision of an online training course, and the implementation of a feedback survey among all NEPS data users.

Moreover, we are active in fostering the visibility of our panel data. Besides activities such as data workshops, conference booths, and producing extensive documentation material for researchers, we are interested in increasing the visibility of research data per se. In Germany, there are initiatives to bring together metadata of relevant panel studies and—in this way—to give researchers an easy way to find the right database for his or her research interests (see also the example of Closer in the United Kingdom[4]). Moreover, we are working on improving the use of our data by the scientific community. Citation of data used in scientific analyses should be as self-evident as the citation of literature references. Until now, clear data citation has not been established as common practice in the educational sciences and many other research areas so that data are often not as visible as they could and should be. We want to facilitate data citation by improving the motivation of researchers using such databases, and we want to increase data visibility by developing further indicators of data use. Concerning the latter, our aim is to analyze whether and to what extent references to data sets on social platforms on the internet or measures for the reception of findings based on the data by the general public in the media may supplement traditional, citation-based impact factors. Improving data citation habits and data visibility is a big challenge—however, it is essential for all infrastructure facilities collecting and disseminating research data to the scientific community.

References

Artelt, C., Weinert, S., & Carstensen, C. H. (Eds.). (2013). Assessing competencies across the lifespan within the German National Educational Panel Study (NEPS). *Journal for Educational Research Online (JERO), 5*(2), 5-14.

Barkow, I., Leopold, T., Raab, M., Schiller, D., Wenzig, K., Blossfeld, H.-P., & Rittberger, M. (2011). RemoteNEPS: Data dissemination in a collaborative workspace. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 315-325). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

[4] http://www.closer.ac.uk/

Bäumer, T., Preis, N., Roßbach, H.-G., Stecher, L., & Klieme, E. (2011). Education processes in life-course-specific learning environments. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 87-101). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Berlin declaration on open access to knowledge in the Sciences and Humanities. (2003). Retrieved from http://openaccess.mpg.de/67605/berlin_declaration_engl.pdf

Bildungsbericht Deutschland [German National Report on Education]. (2014). *Participation in education* [Data file, table B4-1A]. Retrieved from http://www.bildungsbericht.de/zeigen.html?seite=11127

Blossfeld, H.-P. (1985). *Bildungsexpansion und Berufschancen: Empirische Analysen zur Lage der Berufsanfänger in der Bundesrepublik* [Educational expansion and occupational chances]. Frankfurt, Germany: Campus Verlag.

Blossfeld, H.-P., Roßbach, H.-G., & von Maurice, J. (Eds.). (2011). Education as a lifelong process: The German National Educational Panel Study (NEPS). *Zeitschrift für Erziehungswissenschaft, 14.* Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Blossfeld, H.-P., & von Maurice, J. (2011). Education as a lifelong process. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 19-34). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Bundesministerium für Bildung und Forschung [German Federal Ministry of Education and Research]. (2015, March 8). *Information infrastructure.* Retrieved from http://www.bmbf.de/en/21568.php

Bundesagentur für Arbeit [Federal Employment Agency]. (2012, July). *Arbeitsmarkt in Deutschland. Zeitreihen bis 2011* [The German labor market. Time series until 2011]. Retrieved from http://statistik.arbeitsagentur.de/Statischer-Content/Statistische-Analysen/Analytikreports/Zentrale-Analytikreports/Jaehrliche-Analytikreports/Generische-Publikationen/Arbeitsmarkt-Deutschland-Zeitreihen/Analyse-Arbeitsmarkt-Deutschland-Zeitreihen-2011.pdf

Bundesagentur für Arbeit [Federal Employment Agency]. (2015). *Der Arbeits- und Ausbildungsmarkt in Deutschland–Monatsbericht Februar 2015* [Labor and vocational training market in Germany–Monthly report of February 2015]. Retrieved from http://statistik.arbeitsagentur.de/Statischer-Content/Arbeitsmarktberichte/Monatsbericht-Arbeits-Ausbildungsmarkt-Deutschland/Monatsberichte/Generische-Publikationen/Monatsbericht-201502.pdf

Central Intelligence Agency. (2014). *The world factbook: Population pyramid, 2014* [Demographic Map]. Retrieved from https://www.cia.gov/library/publications/the-world-factbook/geos/gm.html

Deutsche Forschungsgemeinschaft [German Research Foundation]. (2013). *Vorschläge zur Sicherung guter wissenschaftlicher Praxis. Denkschrift. Empfehlungen der Kommission "Selbstkontrolle in der Wissenschaft"* [Proposals for safeguarding good scientific practice. Memorandum. Recommendations of the Commission on Professional Self-Regulation in Science] (2nd Suppl. Ed.). Retrieved from http://www.dfg.de/download/pdf/dfg_im_profil/reden_stellungnahmen/download/empfehlung_wiss_praxis_1310.pdf

Dex, S., & Joshi, H. (2004). *Millennium Cohort Study first survey: A user's guide to initial findings.* Retrieved from http://eprints.ioe.ac.uk/5933/1/MCS1_A_Users_ Guide_To_Initial_ Findings.pdf

Erikson, R., Goldthorpe, J. H., & Portocarero, L. (1979). Intergenerational class mobility in three Western European societies: England, France and Sweden. *British Journal of Sociology, 30*, 341-415.

Ganzeboom, H. B. G., De Graaf, P. M., & Treiman, D. J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*, 1-56.

Ganzeboom, H. B. G. (2010). *Occupational status measures for the new International Standard Classification of Occupations ISCO-08; with a discussion of the new classification.* Retrieved from http://www.harryganzeboom.nl/isol/isol2010c2-ganzeboom.pdf

Gross, C., Jobst, A., Jungbauer-Gans, M., & Schwarze, J. (2011). Educational returns over the life course. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 139-153). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

ILO. (2013). *Correspondence table ISCO-08–ISCO-88 (Excel).* Retrieved from http://www.ilo.org/public/english/bureau/stat/isco/docs/corrtab08-88.xls

Kristen, C., Edele, A., Kalter, F., Kogan, I., Schulz, B., Stanat, P., & Will, G. (2011). The education of migrants and their children across the life course. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 121-137). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

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

Meixner, S., Schiller, D., von Maurice, J., & Engelhardt-Wölfler, H. (2011). Data protection issues in the National Educational Panel Study. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 301-313). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

OECD. (2001). *Knowledge and skills for life. First results from the OECD Programme for International Student Assessment (PISA) 2000.* Retrieved from http://www.oecd.org/education/school/programmeforinternationalstudent assessmentpisa/33691620.pdf

OECD. (2007). *Executive Summary PISA 2006: Science competencies for tomorrow's world.* Retrieved from http://www.oecd.org/pisa/pisaproducts/39725224.pdf

OECD. (2014). *PISA 2012 results in focus: What 15-year-olds know and what they can do with what they know.* Retrieved from http://www.oecd.org/pisa/keyfindings/ pisa-2012-results-overview.pdf

Olczyk M., Will, G., & Kristen, C. (2014). *Immigrants in the NEPS: Identifying generation status and group of origin* (NEPS Working Paper No. 41a). Bamberg, Germany: Leibniz Institute for Educational Trajectories, National Educational Panel Study. Retrieved from https://www.neps-data.de/Portals/0/Working%20Papers/WP_XXXXIa.pdf

Paulus, W., Schweizer, R., & Wiemer, S. (2010). *Klassifikation der Berufe 2010. Entwicklung und Ergebnis* [Classification of Occupations 2010. Development and Result]. Retrieved from http://statistik.arbeitsagentur.de/Statischer-Content/Grundlagen/Methodenberichte/Arbeitsmarktstatistik/Generische-Publikationen/Methodenbericht-Klassifikation-Berufe-2010.pdf

RatSWD [German Data Forum] (2010). *Criteria of the German Data Forum (RatSWD) for the Establishment of Research Data Infrastructure.* Retrieved from http://www.ratswd.de/download/publikationen_rat/RatSWD_FDZCriteria.pdf

RatSWD [German Data Forum] (2015). *Tasks and structures*. Retrieved from http://www.ratswd.de/en/ratswd/tasks-structures

Stanat, P., Artelt C., Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., … Weiß, M. (2002). *PISA 2000. Overview of the study. Design, method and results.* Retrieved from https://www.mpib-berlin.mpg.de/Pisa/PISA-2000_Overview.pdf

Stocké, V., Blossfeld, H.-P., Hoenig, K., & Sixt, M. (2011). Social inequality and educational decisions in the life course. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 103-119). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Treiman, D. J. (1977). *Occupational prestige in comparative perspective.* New York, NY: Academic Press.

Wegener, B. (1985). Gibt es Sozialprestige? [Does social prestige exist?]. *Zeitschrift für Soziologie, 14*, 209-235.

Weinert, S., Artelt, C., Prenzel, M., Senkbeil, M., Ehmke, T., & Carstensen, C. H. (2011). Development of competencies across the life span. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS)* (pp. 67-86). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.

Wenzig, K. (2012). *NEPS-Daten mit DOIs referenzieren* [Citing NEPS data with DOI] (RatSWD Working Paper Series No. 202). Berlin: Rat für Sozial- und Wirtschaftsdaten. Retrieved from http://www.ratswd.de/download/ RatSWD_WP_2012/RatSWD_WP_202.pdf

Wenzig, K. (2013). *Die Metadatenstruktur des Nationalen Bildungspanels* [The Metadata Structure of the National Educational Panel Study] (Discussion Papers, No. 17). Bamberg, Germany: Professorship of Demography, University of Bamberg. Retrieved from https://www.uni-bamberg.de/fileadmin/uni/fakultaeten/sowi_lehrstuehle/bevoelkerungswissenschaft/Publikationen/DiscussionPaper/DP_17- 2013.pdf

Wissenschaftsrat [German Council of Science and Humanities]. (2011). *Recommendations on research infrastructures in Humanities and Social Sciences.* Retrieved from http://www.wissenschaftsrat.de/download/archiv/10465-11_engl.pdf

Wohlkinger, F., Ditton, H., von Maurice, J., Haugwitz, M., & Blossfeld, H.-P. (2011). Motivational concepts and personality aspects across the life course. In H.-P. Blossfeld, H.-G. Roßbach, & J. von Maurice (Eds.), *Zeitschrift für Erziehungswissenschaft, 14. Education as a lifelong process: The German National Educational Panel Study (NEPS).* (pp. 155-168). Wiesbaden, Germany: VS Verlag für Sozialwissenschaften.