
Data Dissemination, Documentation, and User Support

Jan Skopek, Knut Wenzig, Daniel Bela, Tobias Koberg,
Manuel Munz and Daniel Fuß

Abstract

A major goal of the NEPS is to provide scientific use data to the international research community. For this purpose, the NEPS has set up a Research Data Center (RDC) that offers a comprehensive portfolio of services, allowing researchers to gain access to and work with NEPS data effectively. The RDC's support concept combines well-known approaches with innovative means of data documentation, data dissemination, and user support. Important building blocks of our dissemination strategy include the provision of (a) user-friendly and edited scientific use data, (b) flexible data access to the scientific community, (c) sufficient, easy-to-obtain, and clearly arranged documentation of NEPS surveys and data, and (d) extensive user support fostering good scientific practices and high-quality educational research. To achieve the highest standards in publishing panel data, the RDC has established a powerful infrastructure for data management, data dissemination, data documentation, and user support. In this chapter, we discuss the core elements of our concept.

1 Introduction

Implementing a multi-cohort sequence design, the National Educational Panel Study (NEPS) maintains six simultaneous panel studies (starting cohorts) for collecting comprehensive longitudinal data on educational careers and competence development representative of Germany. More than 60,000 target persons as well as additional context persons, such as parents, teachers, and school principals, take part in this large-scale social science endeavor. The NEPS approaches its starting cohorts with at least one wave per year. Moreover, the NEPS has conducted additional state-specific studies in Baden-Württemberg and Thuringia.

A major goal of the NEPS is to provide scientific use data to the international research community. For this purpose, the NEPS set up a Research Data Center (RDC) that offers a comprehensive portfolio of services, allowing researchers to gain access to and work with the NEPS data effectively. The RDC's support concept combines well-known approaches with innovative means of data documentation, data dissemination, and user support. Important building blocks of our dissemination strategy include the provision of (a) user-friendly and edited scientific use data, (b) flexible data access to the scientific community, (c) sufficient, easy-to-obtain, and clearly arranged documentation on the NEPS surveys and data, and (d) extensive user support fostering good scientific practices and high-quality educational research. To achieve the highest standards in publishing panel data, the RDC has established a powerful infrastructure for data management, data dissemination, data documentation, and user support. In this chapter, we discuss the core elements of our concept.

2 Preparation of Scientific Use Files

Consistent with the multi-cohort sequence design of the NEPS, the RDC releases six separate lines of Scientific Use Files (SUF). We expand these lines with new data in the course of annual panel waves. Two additional Scientific Use Files are published for the school-reform studies that are conducted in Thuringia and in Baden-Württemberg. In sum, the NEPS publishes 8 Scientific Use Files that provide comprehensive longitudinal data as a result of the first 5 years of the project. In the following section, we describe some important aspects that are relevant to the publication of Scientific Use Files.

Release Strategy

We aim to provide scientific use data to the scientific community no later than 18 months after the end of fieldwork. Thus, we publish data in the form of releases identified by a unique version number with three digits. As a result, researchers working with NEPS data are able to refer precisely to a specific version of the data. We even preserve and archive older versions of data, a measure that ensures that research can replicate results. Furthermore, the version number also discloses the number of included panel waves as well as major and minor data updates.

Citable Data Using DOIs

Traceability of the research process is a significant issue for ensuring good scientific practice. In this regard, the need for citable datasets has been increasing in recent years. Therefore, for each data release, we assign a unique digital object identifier (DOI) that is registered at *da|ra*¹ (Wenzig, 2012). Using the DOI enables researchers

1 *da|ra* is the registration agency for social and economic data run by GESIS and ZBW. See <http://www.da-ra.de/en/home/>.

to cite NEPS data in a very easy and precise way. The DOI also refers to a landing page at the NEPS web portal that provides basic metadata relating to the data and describing methods of data access.

Scientific Use Data

In order to provide high-quality panel data that are ready for scientific use, thorough preparation and editing of input data are essential. Consequently, before releasing any datasets to the research community, we conduct a set of preparation steps, including anonymization, data cleaning, editing, coding, variable generation, data enrichment, data linkage, and quality testing. Throughout the whole preparation process, the design of user-friendly data structures ensuring a maximum of data usability represents a central concern. Therefore, we work intensively on conceptual questions of how to design convenient data structures of Scientific Use Files. For instance, we offer integrated data structures in long format when reasonable and possible. Furthermore, while keeping the grade of detail high in the original panel data, we additionally generate more user-friendly data files that provide derived variables and data in a rather condensed and analytical form.

Large-Scale Editing of Longitudinal Data

The dissemination of huge collections of empirical data embedded in the complex panel design of the NEPS makes collaborative and systematic preparation of data indispensable (see Bela in this volume). The workflow of data editing mainly incorporates editors from the RDC, but experts in substantial topics across other units in the NEPS also participate in specific coding or edition tasks (e.g., the preparation and scoring of competence data by psychometricians, the coding of countries and languages of origin by experts on migration issues). To achieve high-quality standards in a multi-editor environment, we have built up a collaborative infrastructure and committed all coworkers to principles guiding the data-edition process. In addition to leaving raw data unchanged and organizing the edition process in intermediate steps, a fundamental principle is keeping the data edition traceable and replicable. We strive to achieve this goal through completely syntax-based procedures using the software Stata² as a standard technology. A distributed version-control program enables us to keep track of who changed what in syntax files—a technique commonly used in decentralized software development. As a result, the whole process of preparing Scientific Use Files remains traceable at any point in time because it is documented in high precision. For instance, the release of the first SUF for Starting Cohort 6 (NEPS SC6:1.0.0) represented a major collaborative effort. Up to nine data editors participated in the data preparation process. In sum, editors wrote more than 150,000 lines of syntax code and documented almost 900 code revisions.

2 <http://www.stata.com/>

Integrated Datasets

Due to annual or semiannual data collection sweeps along the different cohorts, the handling of panel-wave data (like data merging or harmonization of variables) would soon turn into a serious and exhausting task for researchers, even for a small number of waves. Thus, in order to minimize the research users' data-management efforts that are necessary to work with NEPS data, we produce integrated longitudinal datasets in long format where reasonable. This also includes the integration and harmonization of the data structure across different waves. For example, we usually harmonize variables over waves and cohorts. In particular, life-course data (episodes and spells) collected retro- and prospectively over waves are integrated and harmonized, which is a complex and time-consuming task. As a result, we significantly reduce the burden of data management on the researchers' side. Beyond that, we provide integrated tracking data for each cohort (via files called "*CohortProfile*") that allow for tracking the respondents' participation in the panel samples over waves.

Coding of Occupations, Branches, Courses, and Regions

The coding of open inputs in questionnaires is crucial for improving the quality and utility of data. A routine task in data editing is the recoding of textual responses to residual categories. However, substantial effort goes into the coding of occupations, which requires a significant amount of occupational knowledge and methodological know-how. The coding is mostly done manually, but automatic procedures (e. g., string comparison with keyword lists) that support the coding process have also been developed (see Munz, Wenzig, & Bela in this volume). We apply a set of quality measures in order to optimize the coding outcomes.

Occupations and vocational trainings are coded into the German classification of occupations ("Klassifikation der Berufe," KldB-2010, cf. Paulus, Schweitzer, & Wiemer, 2010). Furthermore, occupational codes according to the KldB-88 and the International Standard Classification of Occupations (ISCO; cf. ILO 2008) from 1988 (ISCO-88) as well as from 2008 (ISCO-08) are provided. Branches are coded according to the classification of branches by official statistics ("Klassifikation der Wirtschaftszweige," WZ08, cf. DESTATIS, 2008). The content of further education courses, such as those collected in Starting Cohort 6, are coded according to the catalog of competencies provided by the Federal Employment Agency ("Kompetenzkatalog der Bundesagentur"), and course identification numbers ("Kurskennziffern") of the catalog are assigned. Moreover, data on locations (like the location of employment or vocational trainings, current place of residence, place of birth) are coded to districts ("Kreise") using official district numbers ("Kreiskennziffer").

Scales and Generated Variables

To improve the usability of datasets, we generate useful additional variables. We deliver additional variables that measure socioeconomic status and occupational prestige, such as 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, & de Leeuw, 1992; Ganzeboom, 2010), the EGP classes (Erikson, Goldthorpe, & Proto-carero, 1979), and the occupational class scheme from Blossfeld (Blossfeld, 1985; Schimpl-Neimanns, 2003) by default. To classify educational attainment, we generated variables for CASMIN and ISCED-97. Relying on the threefold NUTS³ hierarchy of regional clusters, we provide variables containing administrative regions (“Regierungsbezirke,” NUTS Level 2) and federal states (“Bundesländer,” NUTS Level 1) from the coded districts (NUTS Level 3).⁴ Several other variables are generated that measure more specific issues, such as migration background, or that are needed for technical reasons.

Generated Files

We also provide generated data files that offer the user more simple data structures. These files consist of generated variables to a large extent. In particular, in enhancing the usability of the life-course data of Starting Cohort 6, we provide complex life-course data of an adult population in an easy-to-understand and condensed representation. For instance, we generated a data file titled “*Education*,” which provides simple-to-use data on educational transitions across an individual’s life course already coded in CASMIN and ISCED-97 (see Skopek & Munz in this volume). Another generated file contains all transitions of individuals in marital status reconstructed from comprehensive data on partnership biographies. Further biographical information are compiled in a user-friendly way in specific datasets on “*Further Education*,” “*Vocational Training*,” “*Employment*,” “*Military/Civilian Service*,” “*Children*,” and so on. We plan to extend the provision of generated files in the future in collaboration with researchers. In principle, users of the NEPS could also develop and provide generated files that might become published on the website or even part of the Scientific Use File in future data versions.

Weights

The method group of the NEPS prepares three basic types of weights that enable enhanced data analyses. For each first wave data release, (1) design weights, (2) nonresponse adjusted weights, and, if appropriate, (3) post-stratification weights are available. In particular, design weights are important since they account for the unequal selection probabilities in the sampling. A complete revision of implemented sampling strategies can be found in Aßmann et al. (2011). For some starting cohorts, additional replication weights complete the portfolio (Zinn, 2013).

3 NUTS stands for “nomenclature des unités territoriales statistiques,” that is, territorial units for statistics.

4 Currently, we are not allowed to provide NUTS Level 2 or Level 3 variables in conjunction with NEPS data collected in the context of schools or higher-education institutions. However, a federal state identifier (NUTS Level 3) will be available soon for these starting cohorts.

Imputations

Enriching the analytical potential of longitudinal data, the NEPS provides multiple imputations for selected variables containing missing data. Currently, researchers can rely on imputations for income in the context of Starting Cohort 6. The preparation of further files containing plausible values from multiple imputation models (MI files) is planned for the future.

Record Linkage With Administrative Data

Linkage of data—particularly of administrative data—is an important source for improving the quality of survey data. Record linkage with process data is a multifaceted venture in methodological, technical, and data-protection terms. However, conceptual and technical questions have already been clarified together with experts from the Institute for Employment Research (IAB) in Nuremberg. Data are linked on the basis of the respondents' explicit consent. The linkage is established at an individual level by means of probability matching relying on address data as well as on individuals' basic socio-demographic traits. The NEPS has already achieved a linkage of administrative data from the Federal Employment Agency (Bundesagentur für Arbeit) with the data of all published waves of Starting Cohort 6.

Regional data

The NEPS provides regional and macro-level data that can be merged easily with panel data. Fine-grained regional data up to street-section level are available for all cohorts (data are linked at the address level). Regarding these high-resolution regional data, we offer databases from two leading commercial providers of geodata in Germany: *infas Geodaten (now Nexiga)*⁵ and *microm*.⁶ Additionally, based on demand, we provide the service of matching users' own data on regional indicators (e.g., from the German Federal Statistical Office) with NEPS data. For data-protection reasons, analyses relying on linked regional data are restricted to on-site or remote data usage.

Metadata Enrichment

We utilize a database containing structured metadata (see below) to add rich metadata to the Scientific Use Files. Variable and value labels are added, edited, and checked for correct assignment. We translate these metadata into English, allowing international researchers to work with NEPS data comfortably. We additionally extend data files in Stata format by attaching the corresponding question in the survey instrument to the variables. Using a special Stata command, *infoquery* (Bela, 2013)—which is part of the Stata toolset *NEPStools* provided online by the RDC—users can immediately check how questions of certain variables were phrased (in German as well as translated to English) on the Stata's output console. Although this innovative feature

5 infas Geodaten/Nexiga, see <http://www.nexiga.com/>.

6 microm consumer marketing, see <http://www.microm-online.de/>.

is restricted to Stata, it has turned out to be very helpful for a broad range of users because it brings rich codebook information directly into the data. The same is true for *nepsmiss*, another Stata tool that automatically recodes all of the numeric missing values from the NEPS SUFs (-97, -98, etc.) into Stata's extended missing codes (.a, .b, etc.).

3 Data Dissemination

To grant access to the NEPS data, we have established a threefold infrastructure for data access. Research data is distributed by (a) secure download from the NEPS website, (b) an innovative remote-access technology (*RemoteNEPS*), and (c) on-site access. We designed these modes of data access not only to support the full range of users' interests and to maximize data utility, but also to comply with high standards of data confidentiality. While all three access modes provide scientific use data with a common data structure, they differ with regard to their degree of data anonymization.

Disclosure Control and Data Anonymization

To ensure the best possible confidentiality protection of individuals and individual microdata, the NEPS complies with strict national and international standards. First, the disseminated data have to be de facto anonymous data. This implies that we coarsen or cut off identifiable information to minimize the risk of statistical disclosure. String variables relating to openly asked questions are thoroughly checked. Second, the use of data is strictly confidential and for statistical purposes only. Therefore, access to NEPS data is granted exclusively to researchers (i. e., members of the scientific community) who sign a contract with the NEPS. The NEPS has made a large effort regarding legal regulations to keep the data with as much explanatory power as possible. If data modification is necessary, we only employ non-perturbative methods. Our concept of data dissemination distinguishes between three hierarchical levels of data sensitivity. While having the same dataset structure, data files available "on-site" provide more sensitive information than files available with "remote access," which, in turn, contain more information than the "download" versions of data files.

Data Contracts and User Management

Accessing data presupposes a signed contract that contains several data-use agreements. These stipulations require that the applicant handle the data in a secure and confidential manner. In particular, applicants commit themselves to strict data-protection guidelines that forbid any attempt at re-identification, passing along any data without permission, or using the data for purposes other than the specified research objective. The contract defines serious penalties if these stipulations are violated (e. g., high monetary penalty, proscription, exclusion from further data usage). Only re-

searchers who are members of a scientific institution (university or research institute) are eligible applicants. Researchers must provide a brief description of their project based on NEPS data and have to specify the expected duration of usage as well as further participants in the project. Since the NEPS grants data access for scientific purposes only, we check contract proposals strictly for scientific intentions. Approved data-use agreements are published on the NEPS's homepage together with the projects' description provided by the researchers. Contract documents are available in German as well as in English and are freely accessible on the NEPS website. Finally, after a contract has been approved by the NEPS, the researcher receives an NEPS login consisting of a username and password combination.

Secure Data Download

Researchers with a valid contract are able to download all available Scientific Use Files from the NEPS homepage via a secure SSL connection after login. A lot of additional documentation material necessary for using the datasets is provided.

RemoteNEPS: Access to Data in an Innovative and Secure Research Environment

The data access option *RemoteNEPS* represents a real remote infrastructure that was established by the RDC in addition to well-known on-site data usage and physical distribution of SUF (see Skopek, Koberg, & Blossfeld in this volume). *RemoteNEPS* provides safe and powerful access to sensitive NEPS data in an online research environment equipped with common statistical software packages and tools. Researchers can use *RemoteNEPS* with their NEPS login and an additional biometrical authentication.

On-Site: Acquiring the Greatest Detail of Data in a Physically Controlled Environment

The analysis of highly sensitive microdata is only possible on-site in a controlled physical environment at the Leibniz Institute for Educational Trajectories in Bamberg. On-site usage pertains mainly to the analysis of fine-grained regional data in combination with survey data as well as very sensitive items. The secure site prevents any copying or removing of sensitive data from the premises of the NEPS. All input and output devices are locked down, and the computers are not connected to the Internet or any other local area network. RDC staff is allowed to monitor any work performed on the data at all times. Any access to printers is controlled, and outputs process a review before they are provided (output control).

4 Documentation

Comprehensive and accessible documentation is crucial for good scientific research. Thus, the RDC has established a documentation structure relying on an integrated approach to metadata management. English documentation as well as powerful meta-

data services and tools are provided. All information is available on the research-data web portal maintained by the RDC.

Integrated Management of Metadata

A majority of the numerous NEPS substudies usually involve several instruments (i. e., survey questionnaires or competence tests) that typically define dozens of questions and items as well as filtering or interviewer instructions. Many of these items are repeatedly deployed for collecting panel data not only within one cohort, but also over different cohorts. As a consequence, an extraordinary abundance of metadata have to be administered and documented. Additionally, both metadata and the resulting documentation material have to be accessible in the English language since the NEPS aspires to deliver data to an international scientific community of educational researchers. In effect, metadata management is a crucial and nontrivial task at the NEPS.

The NEPS's metadata strategy strives for a structured approach to documentation (see Wenzig et al. in this volume). Only a structured documentation enables us to efficiently link, de-duplicate, reuse, and present all metadata. For this purpose, the RDC has developed a relational SQL database in collaboration with the German Institute for International Educational Research (DIPF) that enables the storing and linking of diverse metadata on studies, instruments, items, datasets, and variables in a systematic, powerful, and highly consistent fashion. As a crucial feature, metadata entities can be cross-referenced, for example, questionnaire items can be linked to datasets, allowing a dynamic documentation that directly leads from a dataset variable to the corresponding question in a questionnaire. In addition, the reuse of metadata enhances data quality because it allows for the tracking of inadvertent changes in variables across panel waves and starting cohorts. The structured documentation also enables an efficient translation of metadata as one has to translate reused elements only once.

As a result of a systematic approach to metadata management, researchers working with NEPS data enjoy a high documentation utility. Since we maintain metadata centrally in a database, corrections and extensions become effective in all derived documentation material, including multilingual codebooks, survey instruments, and dataset labels, in a synchronous and consistent manner. As described above, we even enrich the data files by using the meta-database.

Bilingual Metadata

To facilitate the international use of NEPS data, we consequently translate metadata of survey instruments (e.g., questions, answer schemes) and datasets (variable and value labels) into English. Since the translation of survey instruments is a complex and difficult task, we outsource this to professional translation agencies. We also rely on bilingual metadata for providing bilingual variable- and value labels in datasets.

NEPSplorer: An Efficient Tool for Searching the Meta-Database

In collaboration with the software-development unit at the NEPS and DIPF, the RDC has developed and published the online service *NEPSplorer*, which provides an efficient metadata service that enables the researcher to interactively explore, conveniently search, and quickly retrieve metadata. Like a search engine for the NEPS, it offers a full-text search for all documented metadata of survey instruments and Scientific Use Files. Users can search for and browse any items and variables of interest. For each item, information on question phrases, corresponding variables, answer categories, interview instructions, concepts, keywords, and many other things is available. The tool also displays cross-links between items in survey instruments and variables in Scientific Use Files. Users can store items of interest in a watch list and print an overview of these items. Furthermore, descriptions of surveys and starting cohorts are available. We have optimized the usability of the service by employing a modern asynchronous web frontend that possesses minimal response times.

Data Manuals, Codebooks, and Technical Reports

Apart from the SUF, the RDC prepares enhanced written documentation that is available for download. Most importantly, Scientific Use Files are equipped with *Data Manuals* (in English). The idea behind providing these manuals is to reduce usage hurdles by offering a succinct and user-friendly introduction to the data. For example, our data manuals describe the surveys, the file structure of datasets, content of data files, and the logic of file merging (for an example, see Skopek, 2013). Furthermore, the manuals include exemplary Stata and SPSS syntax that introduce typical data-management operations, such as merging files, handling spell data, and using weights while accounting for sample stratification.

In addition, we provide codebooks and a set of technical reports relating to the Scientific Use Files. The latter include methods reports that document the sampling and fieldwork process, weighting reports, anonymization reports, and data reports. Moreover, there are further supplements, including how-to guides (e.g., working with regional data) and interviewer manuals.

Finally, we also offer so-called *semantic data structure files*. These are data files of a Scientific Use File that have been emptied and thus contain variables and metadata (variable and value labels) but no data rows. These semantic files provided in Stata and SPSS format allow researchers to easily and intuitively explore data files without accessing real data and before signing a data-use agreement with the NEPS.

5 User Support

Currently (as of February 2015), more than 800 researchers have concluded a data contract with the NEPS. Up to 25 new valid contracts arrive per month. Hence, the NEPS is facing a rising demand for data. To facilitate proper usage of the NEPS data,

the RDC offers extensive user support. At its heart, the support program provides training courses including a comprehensive portfolio of theoretical, methodological, and technical topics relevant for working with NEPS data. We hold courses on a regular basis—about 8 to 10 per year—at our training facilities in Bamberg. Moreover, to proliferate the NEPS data, we occasionally provide courses off-site, which mostly take place abroad. User-training sessions usually come as two-day courses. While the first day provides a general overview of the NEPS study, the structure of datasets, the terms and conditions of data usage, and issues of privacy and data protection, the second day includes in-depth presentations, extended exercises, and hands-on data sessions.

To ensure continuous user support, the RDC additionally maintains an email hotline as well as a telephone hotline. The email hotline is supported by an electronic ticket system that facilitates an efficient internal workflow. The phone hotline is available at a separate phone number from Monday to Friday. Our hotline support provides a very high degree of individualized support. Nevertheless, we also provide on-demand, hands-on support in methodological and technical terms for data users (e.g., supporting syntax development, revising syntax, etc.).

6 Conclusion

A major mission of the NEPS is to provide high-quality scientific use data to the international research community. The NEPS has been successful in setting up a Research Data Center (RDC) that is capable of offering a comprehensive portfolio of services, allowing researchers to get access to and work with NEPS data effectively and with minimal constraints. The aim of this chapter was to provide an overview of the powerful infrastructure for data management, data dissemination, data documentation, and user support that has been implemented by the NEPS Research Data Center (RDC). Our strategy's cornerstones embrace the provision of (a) user-friendly and pre-edited scientific use data, (b) flexible data access for scholars, (c) clearly arranged documentation, and (d) comprehensive user support. As a result, a series of highly innovative approaches, instruments, and tools have been developed thanks to a young and highly motivated team that strives to achieve the highest standards in publishing panel data as well as to an adequate funding situation in the NEPS. These ingredients are important for promoting good scientific practices and high-quality educational research.

Finally, it should be noted that even if data management, data dissemination, and user support are crucial issues in social science's survey projects, they are also the issues that are most understated and underestimated in practice. Consequently, history has shown that many data-collection projects face difficulties or even fail to publish a consistent, usable, and well-documented database in a reasonable amount of time. Hence, researchers should be aware of this at the time of proposal writing.

References

- Aßmann, C., Steinhauer, H. W., Kiesl, H., Koch, S., Schönberger, B., Müller-Kuller, A., ... Blossfeld, H.-P. (2011). Sampling designs of the National Educational Panel Study: Challenges and solutions. 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. 51–65). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bela, D. (2013, July). *Provide, enrich and make accessible. Using Stata's capabilities for dissemination NEPS Scientific Use Data*. Presented at the meeting of German Stata Users Group, Potsdam.
- Blossfeld, H.-P. (1985). *Bildungsexpansion und Berufschancen: Empirische Analysen zur Lage der Berufsanfänger in der Bundesrepublik*. Frankfurt, New York: Campus.
- DESTATIS (2008). *Klassifikation der Wirtschaftszweige. Mit Erläuterungen*. Wiesbaden: Statistisches Bundesamt.
- Treiman, D. J. (1977). *Occupational prestige in comparative perspective*. New York: Academic Press.
- 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*(4), pp. 341–415.
- Ganzeboom, H. (2010). *Questions and answers about ISEI-08*. Retrieved from <http://home.fsw.vu.nl/hbg.ganzeboom/isco08/qa-isei-08.htm>
- Ganzeboom, H., de Graaf, P. M., Treiman, D. J., & de Leeuw, J. (1992). A standard international socio-economic index of occupational status. *Social Science Research, 21*(1), pp. 1–56.
- ILO (2008). *International Standard Classification of Occupations. Structure, group definitions and correspondence tables*. Geneva: International Labour Office.
- Skopek, J. (2013). *Starting Cohort 6: Adult Education and Lifelong Learning*. (NEPS Technical Report No. 3.0.1). Retrieved from <https://www.neps-data.de/de-de/datenzentrum/forschungsdaten/startkohorteerwachsene.aspx>
- Paulus, W., Schweitzer, R., & Wiemer, S. (2010). *Klassifikation der Berufe 2010. Entwicklung und Ergebnis* (Methodenbericht der Statistik der Bundesagentur für Arbeit). Nürnberg: Bundesagentur für Arbeit.
- Schimpl-Neimanns, B. (2003). *Umsetzung der Berufsklassifikation von Blossfeld auf die Mikrozensus 1973–1998* (ZUMA-Methodenbericht 2003/10). Mannheim: ZUMA.
- Wegener, B. (1985). Gibt es Sozialprestige? *Zeitschrift für Soziologie, 14*(3), pp. 209–235.
- Wenzig, K. (2012). *NEPS-Daten mit DOIs referenzieren*. (RatSWD Working Paper Series No. 202). Retrieved from <http://EconPapers.repec.org/RePEc:rsrw:rswwps:rswwps202>
- Zinn, S. (2013). *Replication weights for the cohort samples of students in grade 5 and 9 in the National Educational Panel Study*. (NEPS Working Paper No. 27). Bamberg: University of Bamberg, National Educational Panel Study.

About the authors

D. Bela
Leibniz Institute for Educational Trajectories (LifBi),
Bamberg, Germany.

D. Fuß
Leibniz Institute for Educational Trajectories (LifBi),
Bamberg, Germany.

T. Koberg
Leibniz Institute for Educational Trajectories (LifBi),
Bamberg, Germany.

M. Munz
University of Erlangen-Nuremberg,
Erlangen, Germany.

J. Skopek
European University Institute (EUI),
Florence, Italy.
e-mail: Jan.Skopek@EUI.eu

K. Wenzig
German Institute for Economic Research (DIW Berlin),
Berlin, Germany.