

Starting Cohort 4: Grade 9 (SC4)
SUF Version 6.0.0
Anonymization Procedures
Tobias Koberg

Copyrighted Material
Leibniz Institute for Educational Trajectories (LifBi)
Wilhelmsplatz 3, 96047 Bamberg
Director: Prof. Dr. Hans-Günther Roßbach
Executive Director of Research: Dr. Jutta von Maurice
Executive Director of Administration: Dr. Robert Polgar
Bamberg, 2016

1 Preamble

This documentation gives an exhaustive explanation of all disclosure risk minimization techniques applied before dissemination to the Scientific Use File (SUF) of Starting Cohort (SC) SC4, version 6-0-0 (doi:10.5157/NEPS:SC4:6.0.0). For a quick reference to all variables and datafiles which are affected in detail and on which level you will find the full information, please skip forward to appendix section A.1. If you have inquiries or need further information not covered in this document please feel free to contact the research data center at fdz@lifbi.de. Please note the following general remarks:

- in all tables, the anonymization suffix (`_O`, `_R`, `_D`) was removed from the variable name to enhance readability and remove duplicates. To find a specific variable in the data file, add the corresponding suffix or use a wildcard.
- to ensure a data-consistent documentation, all tables in this document have been produced on basis of the SUF data. Variable labels in the data have been cut off at 80 characters, which inherits to this documentation.
- as additional documentation, a *Semantic Data Structure File* is available. This is the complete SUF where all cases have been removed. Feel free to use this as a first overview to the data in cases you do not have access to the real data (e.g. you are interested in data files which are available OnSite only). You are well advised to use this preparing your syntax *before* initiating your travel to our OnSite facility. You find it freely available on the documentation page to this SUF:

www.neps-data.de > Data Center > Data and Documentation > SC4 > Documentation

1.1 Specifications

To ensure the best possible confidentiality protection of individuals and individual micro data, the National Educational Panel Study complies with strict international standards. To achieve this, a *portfolio approach*, based on the ideas of Lane, Heus, and Mulcahy, 2008, has been established. This approach interlocks data security mechanism on five different levels, while the actual data modification (i.e. anonymization) is one of them. For further information, see the oncoming publication Koberg, 2016. Operationalizing this modification procedure, it can be abstracted to the following two criteria:

- the disseminated data was transferred to so called *de facto anonymous data*. Identifiable information is coarsened or cut off and kept securely to minimize the risk of statistical disclosure.
- the use of data is strictly confidential and for statistical purposes only. The closed contract only grants access to members of the scientific community. This contract has a vast amount

1 Preamble

of legal stipulations, one of them being a large fine which applies for the realization of re-identification on purpose. Therefore, the disseminated data is highly protected by law and allows a more flexible range of available data.

To pick up the latter, the NEPS has made a huge effort regarding legal regulations to offer as much analysis power of data as possible. This *paradigm of information esteem* reveals the fact that conducted measures of statistical disclosure control are few. Also, if there really was a need for modification, only non-perturbative methods were used.

1.2 Onion-shaped model

The NEPS grants the user three different modes of data access: (1) *OnSite*, which stands for the opportunity to use the secured infrastructure made available at the LIfBi in Bamberg, (2) *RemoteNEPS*, which is a progressive remote access technology providing a virtual desktop, and finally (3) *Download*, indicating the possibility to fetch data via a secure web portal.

These given access modes have been originated to allow anonymization routines for a subtle differentiation of information. The three resulting levels of anonymization define as follows:

- data provided OnSite is generally not further anonymized. However, even those data was rendered *de facto anonymous*, for no disclosure risk to persist. All information contained remains completely sane. Although users have to deal with limited possibilities of data access (i.e. supervised import and export of their results), they are free to work with all data available in a secure environment.
- access via RemoteNEPS is considered equivalent to OnSite, hence most of the data stays complete.
- as Download is assumed to be the most hazardous access mode¹, some more anonymization techniques are done to the dataset.

Obviously this approach results in three different versions of all involved datasets. To enable a consistent structure, these data files always contain the entire set of variables; it is their content which differs through the three levels.

As normally there is no need to resign aggregated variables in the higher levels (i.e. OnSite or RemoteNEPS), those are already defined as a surplus to the original variable in the OnSite-version. Stepping down to RemoteNEPS, the content of related variables too sensitive for this level is overwritten with an exclusive missing code – an operation which we define as *purging*. Note that system missing values are not affected, allowing the user to differ between value existence and nonexistence. This still is a valuable additional information. The same logic applies to the transition from RemoteNEPS to Download.

¹ 'hazardous' in terms of: the downloaded content is no longer under physical control of the LIfBi

2 Conducted measures

While there is no explicit documentation to this fact, it should remain clear that this procedure accumulates, i.e. purged content under RemoteNEPS is therefore neither included in RemoteNEPS nor in Download.

This *onion-shaped model* provides both ease of (1) use of different sensitivity models (e.g. preparing an analysis using the Download dataset and conducting it afterwards using the OnSite-data) and (2) documentation, for the subject of documentation is the most sensitive level (OnSite), with RemoteNEPS and Download levels being a subset of these data.

The fourth layer *master* depicted below contains every material which is needed during data processing by the NEPS, but is not meant for the scientific community to be usable.

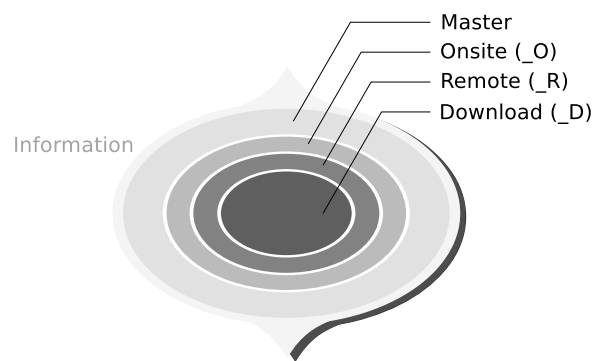


Figure 1: Onion-shaped model defining the different anonymization levels

Technically, this model realizes in a single letter suffixed to dataset and variable names. All datasets available OnSite only are marked with an additional **_O**, those available via RemoteNEPS with **_R** and Download files with **_D**. The same procedure applies when it comes to variable differentiation. A variable which is available unaltered OnSite only is suffixed with **_O**. In RemoteNEPS-access or Download, this variable is still present but purged. If there is an alternate version (e.g. with coarsened content) for RemoteNEPS (suffix **_R**) or Download (suffix **_D**), those can be used. As said before, these are already integrated in the OnSite version. See section 3 for some examples (especially for a practical explanation what is meant by *purging* here).

2 Conducted measures

Keeping the usability and the paradigm of information esteem in mind, only very few alterations are actually done to the dataset. Please note that all information is still available somewhere and that only RemoteNEPS and (mainly) the Download version are constraint in this matter. In fact, in this SUF, only 318 out of 5377 variables are modified in some way.

2 Conducted measures

The following table gives an explanatory overview of all measures conducted. The remainder of this section gives a more detailed description of those topics.

Table 1: Availability of sensitive data

Topic	OnSite	RemoteNEPS	Download
String variables	anonymized	n/a	n/a
International ¹	full data	full data	collapsed
Regional (National)	restricted	restricted	restricted
Institutional context	full data	anonymized	n/a
Macro indicators	accessible	n/a	n/a
Others	full data	full data	aggregated

¹ international geographical information (e.g., nation states, national languages)

2.1 String variables

All variables containing actual text and whose type is therefore string are purged in the RemoteNEPS version. The information remains accessible OnSite. However, you should be aware that in the majority of cases, text entries have been recoded into derived variables (usually suffixed with `_g1`, `_g2`, etc.), which contain the information collected. Those variables are available in all three levels, only the actual text entry is available OnSite only. The text entries have been reviewed by staff to ensure that absolutely no reidentificational material is included. To indicate the availability of those variables OnSite only, the variable names have been suffixed with an additional `_0`.

Table 2: All string variables; those are only available onsite

file	variable	label
pInstitution	h22131x	School: all-day school programs: other, text 1
pInstitution	h22131y	School: all-day school programs: other, text 2
pInstitution	h22131z	School: all-day school programs: other, text 3
pInstitution	h22202t	School: quality: other, text
pInstitution	h227011	School: Teaching staff, amount of teachers per subject, other subjects, text 1
pInstitution	h227012	School: Teaching staff, amount of teachers per subject, other subjects, text 2
pInstitution	h227013	School: Teaching staff, amount of teachers per subject, other subjects, text 3
pInstitution	h229003	School: approach, educational, text
pInstitution	h229005	School: approach, promotion, text

(...)

2 Conducted measures

Table 2: (continued)

file	variable	label
pInstitution	h229007	School: approach, integration, text
pInstitution	h229009	School: approach, other, text
pInstitution	h22901t	School: profile, other, text
pInstitution	h41700a	Other remedial measures, other
pInstitution	h41702a	Parent training programs, migration, other
pInstitution	h41704a	Teacher training programs, migration, other
pParent	p26210t	Parents: tutoring, other subject (open)
pParent	p296402	Wish: What vocation should TC learn
pParent	p401001	Migrant status - other
pParent	p731803	Highest educational qualification of respondent type open
pParent	p731814	Vocational qualification respondent (open)
pParent	p731817	Type tertiary qualification respondent (open)
pParent	p731853	Highest educational qualification partner type open
pParent	p731864	Vocational degree partner (open)
pParent	p731867	Type, tertiary qualification, partner, open
pTarget	t26212a	Student: Subject 1
pTarget	t26212b	Student: Subject 2
pTarget	t26212c	Student: Subject 3
pTarget	t268111	Student: tutoring location somewhere else, text
pTarget	t27111t	Courses outside school: Other courses (open)
pTarget	t27270h	Student: Participation organization/ group: Other, text
pTarget	t31060a	Idealistic vocational aspirations: preferred choice of career
pTarget	t31160a	Realistic vocational aspirations: preferred choice of career
pTarget	t428350	Self-description sense of belonging
pTarget	t516202	Party election (another party)
pTarget	td0026x	Favorite subject
pTarget	te05017	Idealistic educational aspiration - type of higher education institution - other
pTarget	te05025	Idealistic educational aspiration - type of vocational training - something else
pTarget	te06010	Subject 1
pTarget	te06020	Subject combination
pTarget	te08010	Profession requiring vocational training
pTarget	te09017	Realistic educational aspiration - type of higher education institution - other,
pTarget	te09025	Realistic educational aspiration - type of vocational training - something else,
pTarget	te10010	Realistic educational aspiration _subject
pTarget	te10020	In combination with the subject(s)
pTarget	te11010	Realistic educational aspiration - vocational training
pTarget	tf00070	first job wish
pTarget	tf0013b	second vocational wish indication
spCourses	t272011_w1	Course content
spCourses	t272011_w2	Course content

(...)

2 Conducted measures

Table 2: (continued)

file	variable	label
spCourses	t272011_w3	Course content
spParentGap	ps29102	Other activity
spParentSchool	p723090	Type of school (open)
spSchool	ts11206	Another type of school
spSchool	ts11212	Other school-leaving qualification
spSchool	ts11217	Other prospective school-leaving qualification
spSchoolExtExam	ts11305	Other school-leaving certificate awarded
spVocExtExam	ts15305	Other external examination qualification
spVocTrain	ts15220	other vocational qualification
spVocTrain	ts15222	Other qualification (open)

2.2 International: countries, languages, and nationalities

All information corresponding to international localization, nationality or languages is completely available only OnSite or via RemoteNEPS. Variables comprised in the Download SUF are aggregated to the binary information German/non-German. Please note that this variable does not necessarily has the same root of the variable name as the affected variable (e.g. if it is part of the instrument and not generated during the anonymization process). Look in the vicinity of the documented variable to find it. Also, see section 3.3.

Table 3: All variables which are affected by country/language/nationality aggregation

file	variable	label
pParent	p32601a	Position generator: country nurse or male nurse
pParent	p32601b	Position generator: country engineer
pParent	p32601c	Position generator: country warehouse/transport worker
pParent	p32601d	Position generator: country social worker
pParent	p32601e	Position generator: country sales clerk
pParent	p32601f	Position generator: country police officer
pParent	p32601g	Position generator: country doctor
pParent	p32601h	Position generator: country banker
pParent	p32601k	Position generator: country car mechanic
pParent	p32601l	Position generator: country legal practitioner
pParent	p32601m	Position generator: country optician
pParent	p32601n	Position generator: country translator
pParent	p32601o	Position generator: country teacher at elementary school, Hauptschule or Realsch
pParent	p400010_g1	Respondent's country of birth
pParent	p400010_g2	Respondent's country of birth (aggregated)

(...)

2 Conducted measures

Table 3: (continued)

file	variable	label
pParent	p400070_g1	Respondent's mother's country of birth
pParent	p400070_g2	Country of birth of respondent's mother (aggregated)
pParent	p400090_g1	Respondent's father's country of birth
pParent	p400090_g2	Country of birth of respondent's father (aggregated)
pParent	p400500_g3	group of origin
pParent	p401150_g2	Respondent's nationality not German (aggregated)
pParent	p403010_g1	Partner's country of birth
pParent	p403010_g2	Partner's country of birth (aggregated)
pParent	p403070_g1	Partner's mother's country of birth
pParent	p403070_g2	Country of birth of partner's mother (aggregated)
pParent	p403090_g1	Partner's father's country of birth
pParent	p403090_g2	Country of birth of partner's father (aggregated)
pParent	p404050_g2	Other nationality partner (aggregated)
pParent	p406010_g1	Country of birth target child
pParent	p406010_g2	Target child's country of birth (aggregated)
pParent	p407050_g2	Target child's nationality (aggregated)
pParent	p407060_g2	Target child's second nationality (aggregated)
pParent	p731804_g1	Highest educational qualification of respondent abroad
pParent	p731804_g2	Country of respondent's highest educational qualification (aggregated)
pParent	p731823_g1	Country of vocational qualification (additional response)
pParent	p731823_g2	Country of professional qualification (additional response, aggregated)
pParent	p731854_g1	Highest educational qualification partner abroad
pParent	p731854_g2	Country of highest educational qualification partner (aggregated)
pParent	p731873_g1	Country of vocational qualification, partner, (additional response)
pParent	p731873_g2	Country of partner's professional qualification (additional response, aggregated)
pTarget	t400000_g1	Country of own birth
pTarget	t400000_g2	Country of birth (aggregated)
pTarget	t400070_g1	Mother: Country of birth
pTarget	t400070_g2	Mother: Country of birth (aggregated)
pTarget	t400090_g1	Father: Country of birth
pTarget	t400090_g2	Father: Country of birth (aggregated)
pTarget	t400220_g1	Country of birth of your maternal grandparents - maternal grandmother
pTarget	t400220_g2	Mother's mother: Country of birth (aggregated)
pTarget	t400240_g1	Country of birth of your maternal grandparents - maternal grandfather

(...)

2 Conducted measures

Table 3: (continued)

file	variable	label
pTarget	t400240_g2	Mother's father: Country of birth (aggregated)
pTarget	t400260_g1	Country of birth of your paternal grandparents - paternal grandmother
pTarget	t400260_g2	Father's mother: Country of birth (aggregated)
pTarget	t400280_g1	Country of birth of your paternal grandparents - paternal grandfather
pTarget	t400280_g2	Father's father: Country of birth (aggregated)
pTarget	t400500_g3	Group of origin
pTargetCATI	t405010_g1	Birth country
pTargetCATI	t405010_g2	Country of birthplace (aggregated)
pTargetCATI	t406100_g2	other nationality (aggregated)
pTargetCATI	t407010_g1	Partner's country of birth
pTargetCATI	t407010_g2	Country of birth partner (aggregated)
pTargetCATI	t751004_g1	Country of residence (abroad)
pTargetCATI	t751004_g2	Country of place of residence (aggregated)
pTargetCATI	t751015_g1	Country of second residence (outside Germany)
pTargetCATI	t751015_g2	Country of secondary residence (aggregated)
spEmp	ts23239_g1	Location of workplace
spEmp	ts23239_g2	Country of the workplace (aggregated)
spParentSchool	p723060_g1	Country of school
spParentSchool	p723060_g2	Country of school (aggregated)
spSchool	ts1120s_g1	School country
spSchool	ts1120s_g2	Country of school (aggregated)
spSchoolExtExam	ts11301_g1	Country of school-leaving qualification
spSchoolExtExam	ts11301_g2	Country of the awarded school-leaving qualification (aggregated)
spVocExtExam	ts15303_g1	Country of external examination
spVocExtExam	ts15303_g2	Country of external exam (aggregated)
spVocTrain	ts1521s_g1	Country in which vocational training was done
spVocTrain	ts1521s_g2	Country of vocational training (aggregated)
ParentMethods	px80204	Interview: interview language (realized case)
TargetMethods	tx80204	Interview: Interview language (realized case)
pEducator	e41100a_g2	Mother tongue (response 1, ISO 639.2)
pEducator	e41100a_g3	Mother tongue (response 2, ISO 639.2)
pEducator	e41100a_g4	Mother tongue (response 3, ISO 639.2)
pEducator	e41100a_g5	Mother tongue (response 4, ISO 639.2)
pParent	p410000_g1	Child's language of origin (ISO 639.2)
pParent	p410000_g2	Child's language of origin (aggregated)
pParent	p410002_g1	Child's other language of origin (ISO 639.2)
pParent	p410002_g2	Child's other language of origin (aggregated)
pParent	p413000_g1	Respondent's language of origin (ISO 639.2)
pParent	p413000_g2	Respondent's language of origin (aggregated)
pParent	p413002_g1	Respondent's other language of origin (ISO 639.2)
pParent	p413002_g2	Respondent's other language of origin (aggregated)

(...)

2 Conducted measures

Table 3: (continued)

file	variable	label
pParent	p414000_g1	Partner's language of origin (ISO 639.2)
pParent	p414000_g2	Partner's language of origin (aggregated)
pParent	p414002_g1	Partner's other language of origin (ISO 639.2)
pParent	p414002_g2	Partner's other language of origin (aggregated)
pParent	p414100_g1	Recent partner's language of origin (ISO 639.2)
pParent	p414100_g2	Recent partner's language of origin (aggregated)
pParent	p414102_g1	Recent partner's other language of origin (ISO 639.2)
pParent	p414102_g2	Recent partner's other language of origin (aggregated)
pTarget	t41000a_g2	Mother tongue (1st alternative, ISO 639.2)
pTarget	t41000a_g3	Mother tongue (2nd alternative, ISO 639.2)
pTarget	t41000a_g4	Mother tongue (3rd alternative, ISO 639.2)
pTarget	t41000a_g5	Mother tongue (4th alternative, ISO 639.2)
pTarget	t410010_g2	Second language (1st alternative, ISO 639.2)
pTarget	t410010_g3	Second language (2nd alternative, ISO 639.2)
pTarget	t410010_g4	Second language (3rd alternative, ISO 639.2)
pTarget	t410010_g5	Second language (4th alternative, ISO 639.2)
pTarget	t41010a_g2	Mother: Mother tongue (1st alternative, ISO 639.2)
pTarget	t41010a_g3	Mother: Mother tongue (2nd alternative, ISO 639.2)
pTarget	t41010a_g4	Mother: Mother tongue (3rd alternative, ISO 639.2)
pTarget	t41010a_g5	Mother: Mother tongue (4th alternative, ISO 639.2)
pTarget	t41012a_g2	Father: Mother tongue (1st alternative, ISO 639.2)
pTarget	t41012a_g3	Father: Mother tongue (2nd alternative, ISO 639.2)
pTarget	t41012a_g4	Father: Mother tongue (3rd alternative, ISO 639.2)
pTarget	t41012a_g5	Father: Mother tongue (4th alternative, ISO 639.2)
pTargetCATI	t413000_g1	1. language of origin (ISO 639.2)
pTargetCATI	t413000_g2	1st native language (aggregated)
pTargetCATI	t413010_g1	2. language of origin (ISO 639.2)
pTargetCATI	t413010_g2	2nd native language (aggregated)
pParent	p401150_g1	Citizenship respondent not German
pParent	p404050_g1	Other citizenship partner
pParent	p407050_g1	Citizenship of the target child
pParent	p407060_g1	Second citizenship of the target child
pTarget	t40115a_g2	Nationality (response 1; coded)
pTarget	t40115a_g3	Nationality (response 2; coded)
pTarget	t40115a_g4	Nationality (response 3; coded)
pTarget	t42850x_g1	Definition culture of origin
pTargetCATI	t406100_g1	Other citizenship

2.3 Regional localization

For all variables which register a regional position – such as places of birth, work, school, or residence – the following variables have been generated (if possible):

- an indicator for West Germany and East Germany (including Berlin)
- the federal state for this location.

As NEPS data which has been surveyed in institutional context enjoys special focus in this matter, federal states are only available in RemoteNEPS. When using data collected in connection with schools or higher education institutions, it is not allowed to use Federal-State-related information directly or indirectly contained in the data for analyses aiming at a direct Federal State comparison, direct conclusions to be drawn about a Federal State, or a reconstruction of the concrete Federal State affiliation of persons, households, and institutions. Please see the Data Use Agreement for further information about this topic. Regional localization below this level is not directly available. If you have need to merge data using a regional key below those levels, please contact the LifBi RDC (fdz@lifbi.de), which may deal with this matter for you.

Table 4: All regional variables which are only available in RemoteNEPS

file	variable	label
CohortProfile	tx80109_g2	Sample: Federal state school
Weights	stratum_imp2	implicit stratum 2 (federal states of institution)
Weights	stratum_imp3	implicit stratum 3 (regional classification of institution)
Weights	stratum_imp4	implicit stratum 4 (organizing institution)
pEducator	e537110_g2	Place of study teaching post (Federal state)
pEducator	e537110_g3	Place of study teaching post (administrative district)
pEducator	e537110_g4	Place of study teaching post (district)
pEducator	e537170_g2	Place: passed the examination (Federal State)
pEducator	e537170_g3	Place: passed the examination (administrative district)
pEducator	e537170_g4	Place: passed the examination (district)
pParent	p751001_g2	Place of Residence (Federal State)
pTarget	ts11202_g2	Municipality of school (federal state)
pTargetCATI	t700101_g2	Municipality of birth (federal state)
pTargetCATI	t751001_g2	Municipality of Living (RS federal state)
pTargetCATI	t751011_g2	Municipality of secondary residence (federal state)
spEmp	ts23237_g2	Place of workplace (RS federal state)
spParentSchool	p723030_g2	Municipality of school (federal state)
spSchool	ts11202_g2	Municipality of school (federal state)
spVocPrep	ts13105_g2	Municipality of measure (RS federal state)
spVocTrain	ts15207_g2	Municipality of vocational training center (RS federal state)

2 Conducted measures

2.4 Institutional characteristics

In this Starting Cohort, also information about institutions is available. A special focus of anonymization was directed to protection of institutional data, i.e. information about kindergarten, schools, and universities, but also educators and teachers. To minimize the risk of re-identification, this data is only available in RemoteNEPS. This includes the complete datafile(s) `pInstitution`,

but may also affect basic structural details about the group or class (e.g. size of class) in other datafiles. Furthermore, personal information about educators and teachers is treated more securely. You will find detailed information about these subjects from RemoteNEPS onward. Refer to section 2.6 to see the affected variables.

2.5 Macro indicators

Additional information including structural topography and macroeconomic measures was collected. You find those available OnSite only. The following data is available:

- Microm data, describing the surrounding of the targets residence
- Microm data, describing the surrounding of the school
- RegioInfas (infas geodaten) data, describing the surrounding of the targets residence
- RegioInfas data, describing the surrounding of the school

2.6 Other variables

All other variables affected by anonymization and not classifiable in the previous groups are listed below. In the first table (table 5), you find all variables which were aggregated in the Download-SUF. An additional table below each entry displays the values of the variable, both as found in RemoteNEPS (if table column is empty, the variable is numeric and does not have a value label scheme) and in Download.

The second table (table 6) contains all variables which are not aggregated, but purged completely in the Download-version, that is, there is no aggregated version.

2 Conducted measures

Table 5: Other variables which were aggregated in the Download-SUF (aggregation is also stated)

[file] variable: label		
[pCourseClass] e227400_g1: Class: Number of students, total (coarsed)		
RemoteNEPS		Download
	1	Below 10
	2	10 to 14
	3	15 to 19
	4	20 to 24
	5	25 to 29
	6	30 to 34
	7	35 and more
[pCourseClass] e229400: Class: Facilities: Classroom size (aggregated)		
RemoteNEPS		Download
	1	under 30
	2	30 to 39
	3	40 to 49
	4	50 to 59
	5	60 and more
[pEducator] e229820: Work in occupation (years, categorized)		
RemoteNEPS		Download
	1	0 up to below 5
	2	5 up to below 10
	3	10 up to below 15
	4	15 up to below 20
	5	20 up to below 25
	6	25 up to below 30
	7	30 up to below 35
	8	35 and more
[pEducator] e229821: Time in occupation at school (years, categorized)		
RemoteNEPS		Download
	1	0 up to below 5
	2	5 up to below 10
	3	10 up to below 15
	4	15 up to below 20
	5	20 up to below 25
	6	25 up to below 30
	7	30 up to below 35
	8	35 and more
[pEducator] e536020: Age at choice of occupation (categorized)		
RemoteNEPS		Download
	30	30 and older

(...)

2 Conducted measures

Table 5: (continued)

[file] variable: label

[pEducator] e53702y: Year of college admission qualification (categorized)

RemoteNEPS		Download
	1	Before 1970
	2	1970-1979
	3	1980-1989
	4	1990-1999
	5	after 1999

[pEducator] e53710y: Year of study start teacher course (kategorisiert)

RemoteNEPS		Download
	1	Before 1970
	2	1970-1979
	3	1980-1989
	4	1990-1999
	5	after 1999

[pEducator] e537150: Year of passing the examination (categorized)

RemoteNEPS		Download
	1	Before 1970
	2	1970-1979
	3	1980-1989
	4	1990-1999
	5	after 1999

[pEducator] e76212y: Year of birth (categorized)

RemoteNEPS		Download
	1	before 1950
	2	1950-1959
	3	1960-1969
	4	1970-1979
	5	after 1979

(...)

2 Conducted measures

Table 5: (continued)

[file] variable: label

[pParent] p727001: Recommendation secondary school or course of education, which ones? (coarsened)

RemoteNEPS		Download
Hauptschule	1	Hauptschule [basic secondary school]
Realschule	2	Realschule [intermediate secondary school]
Gymnasium	3	Gymnasium [upper secondary school]
Integrated comprehensive school	4	Integrated Gesamtschule [basic and intermediate secondary school]
Combined Hauptschule and Realschule «also Gemeinschaftsschule, Oberschule, Stadtteilschule, extended Realschule, Realschule plus, Werkrealschule»	5	Joint Haupt- and Realschule [basic and intermediate secondary school]
Mittelschule (Saxony, Bavaria)	6	Mittelschule/Regelschule/Sekundarschule/Regionale Schule [basic and intermediate secondary school]
Regelschule (Thüringia)	7	
Regionale Schule (Rhineland-Palatinate)	8	
Sekundarschule (Bremen)	9	
Förderstufe (Hesse)	10	
Special school	11	Special needs school
Orientation stage	12	Orientation stage
School for gifted children	13	School for highly gifted students

[pParent] p727002: Recommendation secondary school or course of education, further ones? (coarsened)

RemoteNEPS		Download
Hauptschule	1	Hauptschule [basic secondary school]
Realschule	2	Realschule [intermediate secondary school]
Gymnasium	3	Gymnasium [upper secondary school]
Integrated comprehensive school	4	Integrated Gesamtschule [basic and intermediate secondary school]
Combined Hauptschule and Realschule «also Gemeinschaftsschule, Oberschule, Stadtteilschule, extended Realschule, Realschule plus, Werkrealschule»	5	Joint Haupt- and Realschule [basic and intermediate secondary school]
Mittelschule (Saxony, Bavaria)	6	Mittelschule/Regelschule/Sekundarschule/Regionale Schule [basic and intermediate secondary school]
Regelschule (Thüringia)	7	
Regionale Schule (Rhineland-Palatinate)	8	
Sekundarschule (Bremen)	9	
Förderstufe (Hesse)	10	
Special school	11	Special needs school
Orientation stage	12	Orientation stage
School for gifted children	13	School for highly gifted students

(...)

2 Conducted measures

Table 5: (continued)

[file] variable: label

[pParent] p731911: Number of employees respondent (aggregated)

RemoteNEPS		Download
1 to 4	1	1 to 4
5 to 9	2	5 to 9
10 to 19	3	10 to 19
20 to 49	4	20 and more
50 to 99	5	
100 to 199	6	
200 to 249	7	
250 to 499	8	
500 to 999	9	
1,000 to 1,999	10	
2,000 and more	11	

[pParent] p731961: Number of employees partner (categorized)

RemoteNEPS		Download
1 to 4	1	1 to 4
5 to 9	2	5 to 9
10 to 19	3	10 to 19
20 to 49	4	20 and more
50 to 99	5	
100 to 199	6	
200 to 249	7	
250 to 499	8	
500 to 999	9	
1,000 to 1,999	10	
2,000 and more	11	

Table 6: Other variables which were purged in the Download-SUF

file	variable	label
pCourseClass	e190011	Class: Amount of students with special needs
pCourseClass	e227400	Class: Number of female students
pCourseClass	e227401	Class: Number of male students
pCourseClass	e451000	Class: Amount of students with migration background
pCourseClass	e79201a	Class: Amount from lower social class
pCourseClass	e79201b	Class: Amount from middle social class
pCourseClass	e79201c	Class: Amount from higher social class
pCourseClass	e79202a	Students with at least one parent with a higher education degree

(...)

3 Examples

Table 6: (continued)

file	variable	label
pEducator	e537030	Federal state higher education entrance qualification (Federal state)
pEducator	e537200	Federal state second state examination (Federal state)
pEducator	e76212m	Month of birth
pInstitution	h76612m	Month of birth
pTarget	t435030_g1	Other Religious Community
pTarget	t43504a_g1	Other religious community (specific)
pTargetCATI	t435030_g1	Religion and religiosity: other religious community

3 Examples

The following examples compare variables from some selected topics. On the left side, you see these variables as they can be found in the RemoteNEPS-SUF (i.e., this is the variant of the variable which contains the full information). On the right side, you find the same variables, but from the Download-version of the SUF, that is, the aggregated or purged variant of the variable. This illustrates the anonymization applied.

3.1 Number of employees

The first example is a variable collected during the interview of Starting Cohort 6 (Adults). The question further investigates the occupational periphery of employers, while here, this employer is the mother of the respondent. On the left, you see the complete information as found in the RemoteNEPS-SUF. This variable is collected using seven classes, where the last four are sparsely filled (variable `t731406_R`, table on the lower left). Because of this, those four codes have been recoded into one category (*»20 and more«*) and a new variable is generated (`t731406_D`, upper left). In the Download-SUF, you find this aggregated variable unaltered (upper right). The variable `t731406_R` rich in content was purged, though. You see that all meaningful codes have been recoded to *»-53 Anonymized«*. Note that the missing code *»-54 Missing by design«* is transported to the Download-SUF unmodified, so you already can derive the actual *statistical power* of this variable.

3 Examples

RemoteNEPS	Download																																																
SC6_pTarget_R_5-1-0.dta (NEPS SUF, SC6 5.1.0 (remote); doi:10.5157/NEPS:SC6:5.1.0)	SC6_pTarget_D_5-1-0.dta (NEPS SUF, SC6 5.1.0 (download); doi:10.5157/NEPS:SC6:5.1.0)																																																
t731406_D -- Number of mother's employees (categorized)	t731406_D -- Number of mother's employees (categorized)																																																
<pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th style="text-align: right;">Freq.</th> </tr> </thead> <tbody> <tr><td>Valid -98 Don't know</td><td style="text-align: right;">7</td></tr> <tr><td>-97 Refused</td><td style="text-align: right;">1</td></tr> <tr><td>-54 Missing by design</td><td style="text-align: right;">36700</td></tr> <tr><td>0 none</td><td style="text-align: right;">423</td></tr> <tr><td>1 1 to 4</td><td style="text-align: right;">330</td></tr> <tr><td>2 5 to 9</td><td style="text-align: right;">64</td></tr> <tr><td>3 10 to 19</td><td style="text-align: right;">22</td></tr> <tr><td>4 20 and more</td><td style="text-align: right;">28</td></tr> <tr><td>Total</td><td style="text-align: right;">37575</td></tr> <tr><td>Missing .</td><td style="text-align: right;">15982</td></tr> <tr><td>Total</td><td style="text-align: right;">53557</td></tr> </tbody> </table> <pre>-----+-----</pre>		Freq.	Valid -98 Don't know	7	-97 Refused	1	-54 Missing by design	36700	0 none	423	1 1 to 4	330	2 5 to 9	64	3 10 to 19	22	4 20 and more	28	Total	37575	Missing .	15982	Total	53557	<pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th style="text-align: right;">Freq.</th> </tr> </thead> <tbody> <tr><td>Valid -98 Don't know</td><td style="text-align: right;">7</td></tr> <tr><td>-97 Refused</td><td style="text-align: right;">1</td></tr> <tr><td>-54 Missing by design</td><td style="text-align: right;">36700</td></tr> <tr><td>0 none</td><td style="text-align: right;">423</td></tr> <tr><td>1 1 to 4</td><td style="text-align: right;">330</td></tr> <tr><td>2 5 to 9</td><td style="text-align: right;">64</td></tr> <tr><td>3 10 to 19</td><td style="text-align: right;">22</td></tr> <tr><td>4 20 and more</td><td style="text-align: right;">28</td></tr> <tr><td>Total</td><td style="text-align: right;">37575</td></tr> <tr><td>Missing .</td><td style="text-align: right;">15982</td></tr> <tr><td>Total</td><td style="text-align: right;">53557</td></tr> </tbody> </table> <pre>-----+-----</pre>		Freq.	Valid -98 Don't know	7	-97 Refused	1	-54 Missing by design	36700	0 none	423	1 1 to 4	330	2 5 to 9	64	3 10 to 19	22	4 20 and more	28	Total	37575	Missing .	15982	Total	53557
	Freq.																																																
Valid -98 Don't know	7																																																
-97 Refused	1																																																
-54 Missing by design	36700																																																
0 none	423																																																
1 1 to 4	330																																																
2 5 to 9	64																																																
3 10 to 19	22																																																
4 20 and more	28																																																
Total	37575																																																
Missing .	15982																																																
Total	53557																																																
	Freq.																																																
Valid -98 Don't know	7																																																
-97 Refused	1																																																
-54 Missing by design	36700																																																
0 none	423																																																
1 1 to 4	330																																																
2 5 to 9	64																																																
3 10 to 19	22																																																
4 20 and more	28																																																
Total	37575																																																
Missing .	15982																																																
Total	53557																																																
t731406_R -- Number of mother's employees	t731406_R -- Number of mother's employees																																																
<pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th style="text-align: right;">Freq.</th> </tr> </thead> <tbody> <tr><td>Valid -98 Don't know</td><td style="text-align: right;">7</td></tr> <tr><td>-97 Refused</td><td style="text-align: right;">1</td></tr> <tr><td>-54 Missing by design</td><td style="text-align: right;">36700</td></tr> <tr><td>0 None</td><td style="text-align: right;">423</td></tr> <tr><td>1 1 to 4</td><td style="text-align: right;">330</td></tr> <tr><td>2 5 to 9</td><td style="text-align: right;">64</td></tr> <tr><td>3 10 to 19</td><td style="text-align: right;">22</td></tr> <tr><td>4 20 to 49</td><td style="text-align: right;">21</td></tr> <tr><td>5 50 to 99</td><td style="text-align: right;">3</td></tr> <tr><td>6 100 to 199</td><td style="text-align: right;">3</td></tr> <tr><td>7 200 to 249</td><td style="text-align: right;">1</td></tr> <tr><td>Total</td><td style="text-align: right;">37575</td></tr> <tr><td>Missing .</td><td style="text-align: right;">15982</td></tr> <tr><td>Total</td><td style="text-align: right;">53557</td></tr> </tbody> </table> <pre>-----+-----</pre>		Freq.	Valid -98 Don't know	7	-97 Refused	1	-54 Missing by design	36700	0 None	423	1 1 to 4	330	2 5 to 9	64	3 10 to 19	22	4 20 to 49	21	5 50 to 99	3	6 100 to 199	3	7 200 to 249	1	Total	37575	Missing .	15982	Total	53557	<pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th style="text-align: right;">Freq.</th> </tr> </thead> <tbody> <tr><td>Valid -54 Missing by design</td><td style="text-align: right;">36700</td></tr> <tr><td>-53 Anonymized</td><td style="text-align: right;">875</td></tr> <tr><td>Total</td><td style="text-align: right;">37575</td></tr> <tr><td>Missing .</td><td style="text-align: right;">15982</td></tr> <tr><td>Total</td><td style="text-align: right;">53557</td></tr> </tbody> </table> <pre>-----+-----</pre>		Freq.	Valid -54 Missing by design	36700	-53 Anonymized	875	Total	37575	Missing .	15982	Total	53557						
	Freq.																																																
Valid -98 Don't know	7																																																
-97 Refused	1																																																
-54 Missing by design	36700																																																
0 None	423																																																
1 1 to 4	330																																																
2 5 to 9	64																																																
3 10 to 19	22																																																
4 20 to 49	21																																																
5 50 to 99	3																																																
6 100 to 199	3																																																
7 200 to 249	1																																																
Total	37575																																																
Missing .	15982																																																
Total	53557																																																
	Freq.																																																
Valid -54 Missing by design	36700																																																
-53 Anonymized	875																																																
Total	37575																																																
Missing .	15982																																																
Total	53557																																																

3.2 Countries (aggregated list)

To illustrate the anonymization of countries, first have a look at this variable from Starting Cohort 6 (Adults). The information («*respondent's country of birth*») was surveyed during the interview. The collected information was recoded during data processing to different derived variables. The first of those is variable `t405010_g1R`, which contains the complete list of countries (lower left, output is shortened to enhance readability). Also, a second variable `t405010_g2` was generated, using a more coarse list. This is helpful if some information was collected, but it is too inaccurate to identify the country (e.g., if someone states he was born in North America, you do not know if he meant U.S. or Canada). This second variable is transported to the Download-SUF unaltered (upper right), which means it in some way represents the anonymized version. The first variable `t405010_g1R` with the full information is purged in the Download-SUF.

3 Examples

RemoteNEPS		Download	
SC6_pTarget_R_5-1-0.dta (NEPS SUF, SC6 5.1.0 (remote); doi:10.5157/NEPS:SC6:5.1.0)		SC6_pTarget_D_5-1-0.dta (NEPS SUF, SC6 5.1.0 (download); doi:10.5157/NEPS:SC6:5.1.0)	
t405010_g2 -- Country of birthplace (categorized)		t405010_g2 -- Country of birthplace (categorized)	
	Freq.		Freq.
Valid		Valid	
-97 Refused	2	-97 Refused	2
-54 Missing by design	36417	-54 Missing by design	36417
-20 Foreign, but not codeable	5	-20 Foreign, but not codeable	5
1 Germany	15388	1 Germany	15388
2 Italy	52	2 Italy	52
3 Poland	244	3 Poland	244
4 Romania	98	4 Romania	98
5 Turkey	182	5 Turkey	182
6 Former Yugoslavia	113	6 Former Yugoslavia	113
7 Former Soviet Union	522	7 Former Soviet Union	522
8 Central and South America, Caribbean	42	8 Central and South America, Caribbean	42
9 Northern and Western Europe	154	9 Northern and Western Europe	154
10 North America	33	10 North America	33
12 Other Middle East and North Africa	73	12 Other Middle East and North Africa	73
13 Other Africa	30	13 Other Africa	30
14 Other Asia	74	14 Other Asia	74
15 Other Central and Eastern Europe	84	15 Other Central and Eastern Europe	84
16 Other Southern Europe	44	16 Other Southern Europe	44
Total	53557	Total	53557
t405010_g1R -- Country of place of birth		t405010_g1R -- Country of place of birth	
	Freq.		Freq.
Valid		Valid	
-97 Refused	2	-54 Missing by design	36417
-54 Missing by design	36417	-53 Anonymized	17140
-20 Foreign, but not codeable	5	Total	53557
0 germany	15388		
120 yugoslavia	30		
121 albania	4		
122 bosnia and herzegovina	22		
124 belgium	5		
:	:		
477 uzbekistan	6		
479 china	7		
482 malaysia	2		
499 other asia	1		
Total	53557		

3.3 Countries (binary variable)

In school cohorts, an aggregation of countries is coarsened to the binary information German/Non-German. To illustrate this, have a look at variables from Starting Cohort 2 (Kindergarten). The information (*»country of birth target child«*) was surveyed during the interview using a two-step procedure: first, the respondent is asked if the child was born in Germany or abroad (variable p406000, top left). This variable is not subject to anonymization and therefore does not have a suffix and is equally existent in both SUF-versions. If the child was born abroad, a second question regarding the specific country followed (variable p406010_g1R, lower left, output is reduced to enhance readability). This second question is purged in the Download-SUF (lower right). There is no need to generate an additional anonymized version of this second question, as the information is already present in the first shown variable (p406000).

3 Examples

RemoteNEPS	Download																																																																
<p>SC2_pParent_R_3-0-0.dta (NEPS SUF, SC2 3.0.0 (remote); doi:10.5157/NEPS:SC2:3.0.0)</p> <p>p406000 -- Target child born in Germany?</p> <pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th>Freq.</th> </tr> </thead> <tbody> <tr> <td>Valid -98 Don't know</td> <td>1</td> </tr> <tr> <td>1 Yes</td> <td>7658</td> </tr> <tr> <td>2 No</td> <td>173</td> </tr> <tr> <td>Total</td> <td>7832</td> </tr> <tr> <td>Missing .</td> <td>3545</td> </tr> <tr> <td>Total</td> <td>11377</td> </tr> </tbody> </table> <pre>-----+-----</pre> <p>p406010_g1R -- Country of birth target child</p> <pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th>Freq.</th> </tr> </thead> <tbody> <tr> <td>Valid -98 Don't know</td> <td>1</td> </tr> <tr> <td>-20 Abroad, but not codeable</td> <td>1</td> </tr> <tr> <td>0 germany</td> <td>7658</td> </tr> <tr> <td>124 belgium</td> <td>2</td> </tr> <tr> <td>125 bulgaria</td> <td>5</td> </tr> <tr> <td>:</td> <td>:</td> </tr> <tr> <td>475 syrian arab republic</td> <td>2</td> </tr> <tr> <td>476 thailand</td> <td>2</td> </tr> <tr> <td>479 china</td> <td>4</td> </tr> <tr> <td>536 new zealand</td> <td>1</td> </tr> <tr> <td>Total</td> <td>7832</td> </tr> <tr> <td>Missing .</td> <td>3545</td> </tr> <tr> <td>Total</td> <td>11377</td> </tr> </tbody> </table> <pre>-----+-----</pre>		Freq.	Valid -98 Don't know	1	1 Yes	7658	2 No	173	Total	7832	Missing .	3545	Total	11377		Freq.	Valid -98 Don't know	1	-20 Abroad, but not codeable	1	0 germany	7658	124 belgium	2	125 bulgaria	5	:	:	475 syrian arab republic	2	476 thailand	2	479 china	4	536 new zealand	1	Total	7832	Missing .	3545	Total	11377	<p>SC2_pParent_D_3-0-0.dta (NEPS SUF, SC2 3.0.0 (download); doi:10.5157/NEPS:SC2:3.0.0)</p> <p>p406000 -- Target child born in Germany?</p> <pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th>Freq.</th> </tr> </thead> <tbody> <tr> <td>Valid -98 Don't know</td> <td>1</td> </tr> <tr> <td>1 Yes</td> <td>7658</td> </tr> <tr> <td>2 No</td> <td>173</td> </tr> <tr> <td>Total</td> <td>7832</td> </tr> <tr> <td>Missing .</td> <td>3545</td> </tr> <tr> <td>Total</td> <td>11377</td> </tr> </tbody> </table> <pre>-----+-----</pre> <p>p406010_g1R -- Country of birth target child</p> <pre>-----+-----</pre> <table border="1"> <thead> <tr> <th></th> <th>Freq.</th> </tr> </thead> <tbody> <tr> <td>Valid -53 Anonymized</td> <td>7832</td> </tr> <tr> <td>Missing .</td> <td>3545</td> </tr> <tr> <td>Total</td> <td>11377</td> </tr> </tbody> </table> <pre>-----+-----</pre>		Freq.	Valid -98 Don't know	1	1 Yes	7658	2 No	173	Total	7832	Missing .	3545	Total	11377		Freq.	Valid -53 Anonymized	7832	Missing .	3545	Total	11377
	Freq.																																																																
Valid -98 Don't know	1																																																																
1 Yes	7658																																																																
2 No	173																																																																
Total	7832																																																																
Missing .	3545																																																																
Total	11377																																																																
	Freq.																																																																
Valid -98 Don't know	1																																																																
-20 Abroad, but not codeable	1																																																																
0 germany	7658																																																																
124 belgium	2																																																																
125 bulgaria	5																																																																
:	:																																																																
475 syrian arab republic	2																																																																
476 thailand	2																																																																
479 china	4																																																																
536 new zealand	1																																																																
Total	7832																																																																
Missing .	3545																																																																
Total	11377																																																																
	Freq.																																																																
Valid -98 Don't know	1																																																																
1 Yes	7658																																																																
2 No	173																																																																
Total	7832																																																																
Missing .	3545																																																																
Total	11377																																																																
	Freq.																																																																
Valid -53 Anonymized	7832																																																																
Missing .	3545																																																																
Total	11377																																																																

3.4 Federal states

Regional locations are usually recoded to different derived variables (see section 2.3). In this example, we focus on the region of the school in Starting Cohort 3 (5th graders). The first derived variable is a binary variable for Western and Eastern Germany (tx80109_g1), present in both SUF-versions (upper left and upper right). The second variable tx80109_g2R, containing the federal states, is only complete in the RemoteNEPS-SUF (lower left), and is purged in the Download-SUF (lower right).

3 Examples

RemoteNEPS		Download	
SC3_CohortProfile_R_3-1-0.dta (NEPS SUF, SC3 3.1.0 (remote); doi:10.5157/NEPS:SC3:3.1.0)		SC3_CohortProfile_D_3-1-0.dta (NEPS SUF, SC3 3.1.0 (download); doi:10.5157/NEPS:SC3:3.1.0)	
tx80109_g1 -- Sample: Federal state school (west/east)		tx80109_g1 -- Sample: Federal state school (west/east)	
	Freq.		Freq.
Valid	-55 Not determinable 458	Valid	-55 Not determinable 458
	1 Western Germany 16299		1 Western Germany 16299
	2 Eastern Germany incl. Berlin 3784		2 Eastern Germany incl. Berlin 3784
	Total 20541		Total 20541
tx80109_g2R -- Sample: Federal state school		tx80109_g2R -- Sample: Federal state school	
	Freq.		Freq.
Valid	-55 Not determinable 458	Valid	-53 Anonymized 20541
	1 Schleswig-Holstein 784		
	2 Hamburg 377		
	3 Lower Saxony 2209		
	4 Bremen 158		
	5 North Rhine-Westphalia 4398		
	6 Hesse 1308		
	7 Rhineland-Palatinate 971		
	8 Baden-Wuerttemberg 3265		
	9 Bavaria 2677		
	10 Saarland 152		
	11 Berlin (Complete) 404		
	12 Brandenburg 1143		
	13 Mecklenburg-Western Pomerania 619		
	14 Saxony 617		
	15 Saxony-Anhalt 401		
	16 Thuringia 600		
	Total 20541		

3.5 Class size

Sometimes, variables are collected numerical. In the example below, the number of students in class was surveyed from the educator in Starting Cohort 2 (Kindergarten) without any further regulations. Thus, the resulting variable e227400_g1R (lower left) does not have a coding scheme. This variable is aggregated to the variable e227400_g1D containing six different class-sizes (upper left). You find this aggregated version in both SUF-versions, the full numerical variable is purged in the Download-SUF, though (lower right).

3 Examples

RemoteNEPS

SC2_pEducator_R_3-0-0.dta
(NEPS SUF, SC2 3.0.0 (remote); doi:10.5157/NEPS:SC2:3.0.0)

e227400_g1D -- Class: number of students, total (coarsed)

	Freq.
Valid -90 Unspecific missing	10
-54 Missing by design	1803
1 Below 10	3
2 10 to 14	26
3 15 to 19	203
4 20 to 24	450
5 25 to 29	169
6 30 to 34	4
Total	2668

e227400_g1R -- Class: number of students total

	Freq.
Valid -90 Unspecific missing	10
-54 Missing by design	1803
8	3
10	1
11	1
12	4
13	8
14	12
15	21
16	22
17	34
18	58
19	68
20	75
21	89
22	98
23	88
24	100
25	83
26	39
27	27
28	14
29	6
30	3
31	1
Total	2668

Download

SC2_pEducator_D_3-0-0.dta
(NEPS SUF, SC2 3.0.0 (download); doi:10.5157/NEPS:SC2:3.0.0)

e227400_g1D -- Class: number of students, total (coarsed)

	Freq.
Valid -90 Unspecific missing	10
-54 Missing by design	1803
1 Below 10	3
2 10 to 14	26
3 15 to 19	203
4 20 to 24	450
5 25 to 29	169
6 30 to 34	4
Total	2668

e227400_g1R -- Class: number of students total

	Freq.
Valid -54 Missing by design	1803
-53 Anonymized	865
Total	2668

A Appendix

A.1 Complete overview

The following table lists all variables which are affected by anonymization methods. For all variables, full information is available OnSite, although for many of them, you already find this complete information in the RemoteNEPS-version. Only for those variables marked with an asterisk, the full information is only present OnSite. For most of the variables, an aggregated or generated version is available, which still transports some information to the Download-version. Please see the previous chapters for more information.

Table 7: Complete overview of all restricted variables

file	variable	label (*only available OnSite)
pParent	p32601a	Position generator: country nurse or male nurse
pParent	p32601b	Position generator: country engineer
pParent	p32601c	Position generator: country warehouse/transport worker
pParent	p32601d	Position generator: country social worker
pParent	p32601e	Position generator: country sales clerk
pParent	p32601f	Position generator: country police officer
pParent	p32601g	Position generator: country doctor
pParent	p32601h	Position generator: country banker
pParent	p32601k	Position generator: country car mechanic
pParent	p32601l	Position generator: country legal practitioner
pParent	p32601m	Position generator: country optician
pParent	p32601n	Position generator: country translator
pParent	p32601o	Position generator: country teacher at elementary school, Hauptschule or Realsch
pParent	p400010_g1	Respondent's country of birth
pParent	p400010_g2	Respondent's country of birth (aggregated)
pParent	p400070_g1	Respondent's mother's country of birth
pParent	p400070_g2	Country of birth of respondent's mother (aggregated)
pParent	p400090_g1	Respondent's father's country of birth
pParent	p400090_g2	Country of birth of respondent's father (aggregated)
pParent	p400500_g3	group of origin
pParent	p401150_g2	Respondent's nationality not German (aggregated)
pParent	p403010_g1	Partner's country of birth
pParent	p403010_g2	Partner's country of birth (aggregated)
pParent	p403070_g1	Partner's mother's country of birth
pParent	p403070_g2	Country of birth of partner's mother (aggregated)
pParent	p403090_g1	Partner's father's country of birth
pParent	p403090_g2	Country of birth of partner's father (aggregated)
pParent	p404050_g2	Other nationality partner (aggregated)
pParent	p406010_g1	Country of birth target child
pParent	p406010_g2	Target child's country of birth (aggregated)
pParent	p407050_g2	Target child's nationality (aggregated)

(...)

Table 7: (continued)

file	variable	label (*only available OnSite)
pParent	p407060_g2	Target child's second nationality (aggregated)
pParent	p731804_g1	Highest educational qualification of respondent abroad
pParent	p731804_g2	Country of respondent's highest educational qualification (aggregated)
pParent	p731823_g1	Country of vocational qualification (additional response)
pParent	p731823_g2	Country of professional qualification (additional response, aggregated)
pParent	p731854_g1	Highest educational qualification partner abroad
pParent	p731854_g2	Country of highest educational qualification partner (aggregated)
pParent	p731873_g1	Country of vocational qualification, partner, (additional response)
pParent	p731873_g2	Country of partner's professional qualification (additional response, aggregated)
pTarget	t400000_g1	Country of own birth
pTarget	t400000_g2	Country of birth (aggregated)
pTarget	t400070_g1	Mother: Country of birth
pTarget	t400070_g2	Mother: Country of birth (aggregated)
pTarget	t400090_g1	Father: Country of birth
pTarget	t400090_g2	Father: Country of birth (aggregated)
pTarget	t400220_g1	Country of birth of your maternal grandparents - maternal grandmother
pTarget	t400220_g2	Mother's mother: Country of birth (aggregated)
pTarget	t400240_g1	Country of birth of your maternal grandparents - maternal grandfather
pTarget	t400240_g2	Mother's father: Country of birth (aggregated)
pTarget	t400260_g1	Country of birth of your paternal grandparents - paternal grandmother
pTarget	t400260_g2	Father's mother: Country of birth (aggregated)
pTarget	t400280_g1	Country of birth of your paternal grandparents - paternal grandfather
pTarget	t400280_g2	Father's father: Country of birth (aggregated)
pTarget	t400500_g3	Group of origin
pTargetCATI	t405010_g1	Birth country
pTargetCATI	t405010_g2	Country of birthplace (aggregated)
pTargetCATI	t406100_g2	other nationality (aggregated)
pTargetCATI	t407010_g1	Partner's country of birth
pTargetCATI	t407010_g2	Country of birth partner (aggregated)
pTargetCATI	t751004_g1	Country of residence (abroad)
pTargetCATI	t751004_g2	Country of place of residence (aggregated)
pTargetCATI	t751015_g1	Country of second residence (outside Germany)
pTargetCATI	t751015_g2	Country of secondary residence (aggregated)
spEmp	ts23239_g1	Location of workplace
spEmp	ts23239_g2	Country of the workplace (aggregated)
spParentSchool	p723060_g1	Country of school
spParentSchool	p723060_g2	Country of school (aggregated)
spSchool	ts1120s_g1	School country
spSchool	ts1120s_g2	Country of school (aggregated)

(...)

Table 7: (continued)

file	variable	label (*only available OnSite)
spSchoolExtExam	ts11301_g1	Country of school-leaving qualification
spSchoolExtExam	ts11301_g2	Country of the awarded school-leaving qualification (aggregated)
spVocExtExam	ts15303_g1	Country of external examination
spVocExtExam	ts15303_g2	Country of external exam (aggregated)
spVocTrain	ts1521s_g1	Country in which vocational training was done
spVocTrain	ts1521s_g2	Country of vocational training (aggregated)
ParentMethods	px80204	Interview: interview language (realized case)
TargetMethods	tx80204	Interview: Interview language (realized case)
pEducator	e41100a_g2	Mother tongue (response 1, ISO 639.2)
pEducator	e41100a_g3	Mother tongue (response 2, ISO 639.2)
pEducator	e41100a_g4	Mother tongue (response 3, ISO 639.2)
pEducator	e41100a_g5	Mother tongue (response 4, ISO 639.2)
pParent	p410000_g1	Child's language of origin (ISO 639.2)
pParent	p410000_g2	Child's language of origin (aggregated)
pParent	p410002_g1	Child's other language of origin (ISO 639.2)
pParent	p410002_g2	Child's other language of origin (aggregated)
pParent	p413000_g1	Respondent's language of origin (ISO 639.2)
pParent	p413000_g2	Respondent's language of origin (aggregated)
pParent	p413002_g1	Respondent's other language of origin (ISO 639.2)
pParent	p413002_g2	Respondent's other language of origin (aggregated)
pParent	p414000_g1	Partner's language of origin (ISO 639.2)
pParent	p414000_g2	Partner's language of origin (aggregated)
pParent	p414002_g1	Partner's other language of origin (ISO 639.2)
pParent	p414002_g2	Partner's other language of origin (aggregated)
pParent	p414100_g1	Recent partner's language of origin (ISO 639.2)
pParent	p414100_g2	Recent partner's language of origin (aggregated)
pParent	p414102_g1	Recent partner's other language of origin (ISO 639.2)
pParent	p414102_g2	Recent partner's other language of origin (aggregated)
pTarget	t41000a_g2	Mother tongue (1st alternative, ISO 639.2)
pTarget	t41000a_g3	Mother tongue (2nd alternative, ISO 639.2)
pTarget	t41000a_g4	Mother tongue (3rd alternative, ISO 639.2)
pTarget	t41000a_g5	Mother tongue (4th alternative, ISO 639.2)
pTarget	t410010_g2	Second language (1st alternative, ISO 639.2)
pTarget	t410010_g3	Second language (2nd alternative, ISO 639.2)
pTarget	t410010_g4	Second language (3rd alternative, ISO 639.2)
pTarget	t410010_g5	Second language (4th alternative, ISO 639.2)
pTarget	t41010a_g2	Mother: Mother tongue (1st alternative, ISO 639.2)
pTarget	t41010a_g3	Mother: Mother tongue (2nd alternative, ISO 639.2)
pTarget	t41010a_g4	Mother: Mother tongue (3rd alternative, ISO 639.2)
pTarget	t41010a_g5	Mother: Mother tongue (4th alternative, ISO 639.2)
pTarget	t41012a_g2	Father: Mother tongue (1st alternative, ISO 639.2)
pTarget	t41012a_g3	Father: Mother tongue (2nd alternative, ISO 639.2)
pTarget	t41012a_g4	Father: Mother tongue (3rd alternative, ISO 639.2)
pTarget	t41012a_g5	Father: Mother tongue (4th alternative, ISO 639.2)
pTargetCATI	t413000_g1	1. language of origin (ISO 639.2)
pTargetCATI	t413000_g2	1st native language (aggregated)
pTargetCATI	t413010_g1	2. language of origin (ISO 639.2)

(...)

Table 7: (continued)

file	variable	label (*only available OnSite)	
pTargetCATI	t413010_g2	2nd native language (aggregated)	
pParent	p401150_g1	Citizenship respondent not German	
pParent	p404050_g1	Other citizenship partner	
pParent	p407050_g1	Citizenship of the target child	
pParent	p407060_g1	Second citizenship of the target child	
pTarget	t40115a_g2	Nationality (response 1; coded)	
pTarget	t40115a_g3	Nationality (response 2; coded)	
pTarget	t40115a_g4	Nationality (response 3; coded)	
pTarget	t42850x_g1	Definition culture of origin	
pTargetCATI	t406100_g1	Other citizenship	
pInstitution	h22131x	School: all-day school programs: other, text 1	*
pInstitution	h22131y	School: all-day school programs: other, text 2	*
pInstitution	h22131z	School: all-day school programs: other, text 3	*
pInstitution	h22202t	School: quality: other, text	*
pInstitution	h227011	School: Teaching staff, amount of teachers per subject, other subjects, text 1	*
pInstitution	h227012	School: Teaching staff, amount of teachers per subject, other subjects, text 2	*
pInstitution	h227013	School: Teaching staff, amount of teachers per subject, other subjects, text 3	*
pInstitution	h229003	School: approach, educational, text	*
pInstitution	h229005	School: approach, promotion, text	*
pInstitution	h229007	School: approach, integration, text	*
pInstitution	h229009	School: approach, other, text	*
pInstitution	h22901t	School: profile, other, text	*
pInstitution	h41700a	Other remedial measures, other	*
pInstitution	h41702a	Parent training programs, migration, other	*
pInstitution	h41704a	Teacher training programs, migration, other	*
pParent	p26210t	Parents: tutoring, other subject (open)	*
pParent	p296402	Wish: What vocation should TC learn	*
pParent	p401001	Migrant status - other	*
pParent	p731803	Highest educational qualification of respondent type open	*
pParent	p731814	Vocational qualification respondent (open)	*
pParent	p731817	Type tertiary qualification respondent (open)	*
pParent	p731853	Highest educational qualification partner type open	*
pParent	p731864	Vocational degree partner (open)	*
pParent	p731867	Type, tertiary qualification, partner, open	*
pTarget	t26212a	Student: Subject 1	*
pTarget	t26212b	Student: Subject 2	*
pTarget	t26212c	Student: Subject 3	*
pTarget	t268111	Student: tutoring location somewhere else, text	*
pTarget	t27111t	Courses outside school: Other courses (open)	*
pTarget	t27270h	Student: Participation organization/ group: Other, text	*
pTarget	t31060a	Idealistic vocational aspirations: preferred choice of career	*
pTarget	t31160a	Realistic vocational aspirations: preferred choice of career	*

(...)

Table 7: (continued)

file	variable	label (*only available OnSite)	
pTarget	t428350	Self-description sense of belonging	*
pTarget	t516202	Party election (another party)	*
pTarget	td0026x	Favorite subject	*
pTarget	te05017	Idealistic educational aspiration - type of higher education institution - other	*
pTarget	te05025	Idealistic educational aspiration - type of vocational training - something else	*
pTarget	te06010	Subject 1	*
pTarget	te06020	Subject combination	*
pTarget	te08010	Profession requiring vocational training	*
pTarget	te09017	Realistic educational aspiration - type of higher education institution - other,	*
pTarget	te09025	Realistic educational aspiration - type of vocational training - something else,	*
pTarget	te10010	Realistic educational aspiration _subject	*
pTarget	te10020	In combination with the subject(s)	*
pTarget	te11010	Realistic educational aspiration - vocational training	*
pTarget	tf00070	first job wish	*
pTarget	tf0013b	second vocational wish indication	*
spCourses	t272011_w1	Course content	*
spCourses	t272011_w2	Course content	*
spCourses	t272011_w3	Course content	*
spParentGap	ps29102	Other activity	*
spParentSchool	p723090	Type of school (open)	*
spSchool	ts11206	Another type of school	*
spSchool	ts11212	Other school-leaving qualification	*
spSchool	ts11217	Other prospective school-leaving qualification	*
spSchoolExtExam	ts11305	Other school-leaving certificate awarded	*
spVocExtExam	ts15305	Other external examination qualification	*
spVocTrain	ts15220	other vocational qualification	*
spVocTrain	ts15222	Other qualification (open)	*
CohortProfile	tx80109_g2	Sample: Federal state school	
Weights	stratum_imp2	implicit stratum 2 (federal states of institution)	
Weights	stratum_imp3	implicit stratum 3 (regional classification of institution)	
Weights	stratum_imp4	implicit stratum 4 (organizing institution)	
pEducator	e537110_g2	Place of study teaching post (Federal state)	
pEducator	e537110_g3	Place of study teaching post (administrative district)	
pEducator	e537110_g4	Place of study teaching post (district)	
pEducator	e537170_g2	Place: passed the examination (Federal State)	
pEducator	e537170_g3	Place: passed the examination (administrative district)	
pEducator	e537170_g4	Place: passed the examination (district)	
pParent	p751001_g2	Place of Residence (Federal State)	
pTarget	ts11202_g2	Municipality of school (federal state)	
pTargetCATI	t700101_g2	Municipality of birth (federal state)	
pTargetCATI	t751001_g2	Municipality of Living (RS federal state)	
pTargetCATI	t751011_g2	Municipality of secondary residence (federal state)	
spEmp	ts23237_g2	Place of workplace (RS federal state)	
spParentSchool	p723030_g2	Municipality of school (federal state)	

(...)

Table 7: (continued)

file	variable	label (*only available OnSite)
spSchool	ts11202_g2	Municipality of school (federal state)
spVocPrep	ts13105_g2	Municipality of measure (RS federal state)
spVocTrain	ts15207_g2	Municipality of vocational training center (RS federal state)
pCourseClass	e227400_g1	Class: Number of students total
pCourseClass	e229400	Class: Set-up: Classroom size
pEducator	e229820	Time in occupation - overall
pEducator	e229821	Time in occupation - school
pEducator	e536020	Age when profession chosen
pEducator	e53702y	Year of university entrance qualification
pEducator	e53710y	Start of teacher training
pEducator	e537150	Year of state examination
pEducator	e76212y	Year of birth
pParent	p727001	Recommendation for a secondary school or form of education, which?
pParent	p727002	Recommendation secondary school or form of education, secondary?
pParent	p731911	Number of employees respondent
pParent	p731961	Number of employees partner
pCourseClass	e190011	Class: Amount of students with special needs
pCourseClass	e227400	Class: Number of female students
pCourseClass	e227401	Class: Number of male students
pCourseClass	e451000	Class: Amount of students with migration background
pCourseClass	e79201a	Class: Amount from lower social class
pCourseClass	e79201b	Class: Amount from middle social class
pCourseClass	e79201c	Class: Amount from higher social class
pCourseClass	e79202a	Students with at least one parent with a higher education degree
pEducator	e537030	Federal state higher education entrance qualification (Federal state)
pEducator	e537200	Federal state second state examination (Federal state)
pEducator	e76212m	Month of birth *
pInstitution	h76612m	Month of birth *
pTarget	t435030_g1	Other Religious Community
pTarget	t43504a_g1	Other religious community (specific)
pTargetCATI	t435030_g1	Religion and religiosity: other religious community
pTargetMicrom		*
pTargetRegioInfas		*
pInstitutionMicrom		*
pInstitutionRegioInfas		*
pInstitution		*

B References

- Hundepool, A., Domingo-Ferrer, J., Franconi, L., Giessing, S., Nordholt, E. S., Spicer, K., & de Wolf, P.-P. (2012). *Statistical Disclosure Control*. Chichester: Wiley.
- Koberg, T. (2016). Disclosing the National Educational Panel Study. In H.-P. Blossfeld, J. von Maurice, M. Bayer, & J. Skopek (Eds.), *Methodological Issues of Longitudinal Surveys. The Example of the National Educational Panel Study (in press)*. Wiesbaden: Springer VS.
- Lane, J., Heus, P., & Mulcahy, T. (2008). Data Access in a Cyber World: Making Use of Cyber-infrastructure. *Transactions on Data Privacy*, 1, 2–16.
- Müller, W., Blien, U., & Knoche, P. (1991). Die faktische Anonymität von Mikrodaten. *Schriftenreihe Forum der Bundesstatistik*, 19.