



# Provide, Enrich and Make Accessible

Using Stata's Capabilities for Disseminating NEPS Scientific Use Data

Daniel Bela

National Educational Panel Study

NEPS Data Center

University of Bamberg

2013 German Stata Users Group Meeting, Potsdam

07jun2013

SPONSORED BY THE



Federal Ministry  
of Education  
and Research

## Agenda

- 1 About NEPS and NEPS Data Center
- 2 NEPS' metadata system
- 3 Accessing enriched metadata in Stata datasets
- 4 Résumé

## National Educational Panel Study

### NEPS in brief:

- conducts surveys and ...
- ... provides data on educational trajectories to the community
- these data currently consist of ...
  - ... six panel cohorts
  - ... two additional (cross-sectional) studies on school reforms

## National Educational Panel Study

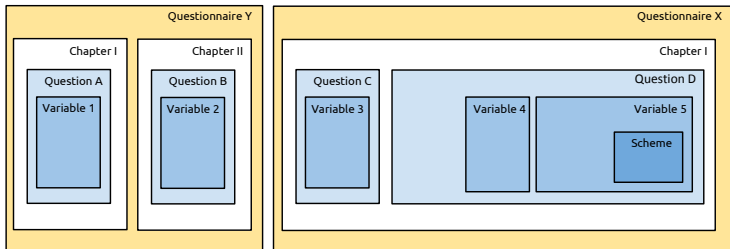
### NEPS in brief:

- conducts surveys and ...
- ... provides data on educational trajectories to the community
- these data currently consist of ...
  - ... six panel cohorts
  - ... two additional (cross-sectional) studies on school reforms

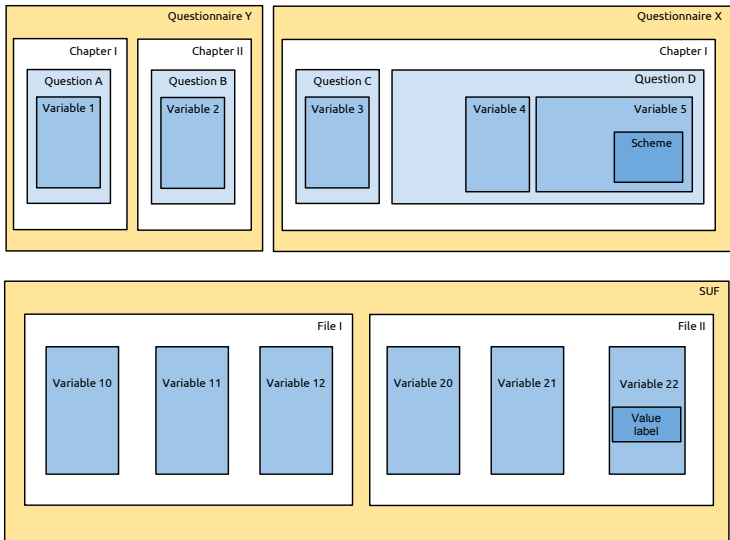
### NEPS Data Center:

- data preparation, integration and edition (using Stata)
- survey and data documentation
- provide data access
- user support

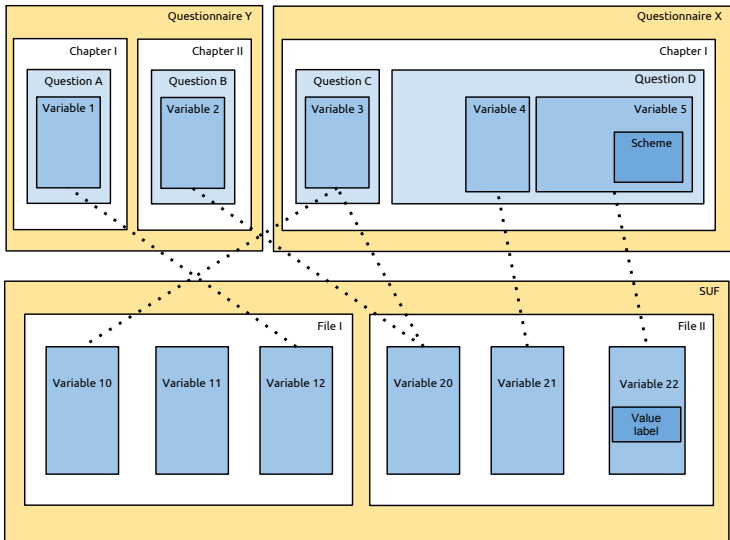
## NEPS' metadata structure (simplified)



## NEPS' metadata structure (simplified)



## NEPS' metadata structure (simplified)



## Incarnations of structured metadata

### Database as “single source of information” for multi-lingual:

- online panel documentation—“NEPSplorer” web application:  
<https://www.neps-data.de/datacenter/studydocumentation/nepsplorer>
- PDF representations of survey questionnaires
- PDF codebooks
- **Stata** (and SPSS) **dataset files**:
  - ① import to Stata via `-odbc load-` and ...
  - ② ... rename variables
  - ③ ... label variables
  - ④ ... define value labels
  - ⑤ ... label values
  - ⑥ ... order variables in datasets
  - ⑦ ... attach additional information as characteristics
  - ⑧ ... suppress variables from dissemination



## Accessibility to the data user : Overview

### Three main ado files:

- infoquery
- charren
- nepsmis

### Installation:

```
net install nepstools, from("http://nocrypt.neps-data.de/stata")
```

## infoquery: Overview

### infoquery.ado: query information attached to variables

- wrapper for `char list`
- presents predefined characteristics of variables
- syntax:

```
infoquery varlist [, alllanguages charlist(charlist)]
```

## infoquery: Example

```
. use "SC4_xTarget_D_1-1-0.dta", clear
```

```
(NEPS SUF, SC4 1.1.0 (download); doi:10.5157/NEPS:SC4:1.1.0)
```

```
. label language en
```

```
. infoquery t514001_w1
```

-----  
query result for variable t514001\_w1:

```
t514001_w1[instname]:  
satis1_w1
```

```
t514001_w1[sufname]:  
t514001_w1
```

```
t514001_w1[questiontext_en]:  
How satisfied are you...
```

```
t514001_w1[variablequestion_en]:  
... with your life at present on the whole?
```

-----

## charren: Overview

### charren: rename variables via characteristics saved in the dataset

- rename variables to predefined names
- different names are saved in characteristics
- reverse search-and-rename possible
- syntax:  
charren `namelist`, to(charname) [`verbose searchspace(charlist)`]

## charren: Example

```
. use "SC4_xTarget_D_1-1-0.dta", clear
```

```
(NEPS SUF, SC4 1.1.0 (download); doi:10.5157/NEPS:SC4:1.1.0)
```

```
. label language en  
. ds t514001_w1 t514002_w1
```

```
t514001_w1 t514002_w1
```

```
. charren t514001_w1 satis2_w1, to(instname)verbose
```

```
Info: will rename t514001_w1 to satis1_w1  
Info: satis2_w1 is not a variable name in current dataset;  
> searching for satis2_w1 in specified search space  
Info: will rename t514002_w1 to satis2_w1
```

```
. ds satis1_w1 satis2_w1
```

```
satis1_w1 satis2_w1
```

## nepsmiss: Motivation

```
. use "SC4_xTarget_D_1-1-0.dta", clear
```

(NEPS SUF, SC4 1.1.0 (download); doi:10.5157/NEPS:SC4:1.1.0)

```
. label language en
. fre t514001_w1, tabulate(4)
```

t514001\_w1 -- Satisfied with life

		Freq.	Percent	Valid	Cum.
Valid	-95 Implausible value	20	0.12	0.12	0.12
	-90 Unspecific missing	195	1.20	1.20	1.32
	-56 Not participated	677	4.17	4.17	5.49
	-54 Missing by design	1087	6.69	6.69	12.18
	:	:	:	:	:
	7 7	2312	14.22	14.22	47.73
	8 8	3566	21.94	21.94	69.67
	9 9	2768	17.03	17.03	86.70
	10 entirely satisfied	2162	13.30	13.30	100.00
	Total	16254	100.00	100.00	

## nepsmiss: Overview

### nepsmiss: create Stata extended missing codes from NEPS missing codes

- (motivation: NEPS delivers missing values as numerical values)
- wrapper for `mvdecode` / `mvencode`
- default list of numeric values
- also rewrites value labels
- simple “reverse” mode
- syntax:  
`nepsmiss varlist [, reverse keeplabels misslist(numlist) verbose]`

## nepsmi: Example

```
. nepsmi t514001_w1
```

Warning: value label de935 is also attached to variables t514002\_w1 [...];  
 > these surplus variables will not be recoded!  
 Consider running nepsmi on those variables as well to correctly recode them!  
 Recoded 1979 values in total

```
. fre t514001_w1, tabulate(1)
```

```
t514001_w1 -- Satisfied with life
```

		Freq.	Percent	Valid	Cum.
Valid	0 entirely dissatisfied	83	0.51	0.58	0.58
	:	:	:	:	:
	10 entirely satisfied	2162	13.30	15.15	100.00
	Total	14275	87.82	100.00	
Missing	.e Implausible value	20	0.12		
	.j Unspecific missing	195	1.20		
	.k Not participated	677	4.17		
	.m Missing by design	1087	6.69		
	Total	1979	12.18		
Total		16254	100.00		



## tl;dr

- 1 NEPS Data Center holds and maintains its questionnaire and dataset metadata in a SQL database
- 2 these data tables are combined to reflect relations between survey and data
- 3 database entries are translated to English
- 4 Stata is used to connect to this database and write (bilingual) information to dataset files
- 5 additional information, such as question texts, is written to `chars`
- 6 ado files are provided for users to access these information
- 7 *data publishers: exploit these features for data documentation!*

## Thank you!

### Questions?

Feel free to ask!

### Interested in NEPS?

- visit <https://www.neps-data.de/>
- write an email to [userservice.neps@uni-bamberg.de](mailto:userservice.neps@uni-bamberg.de)
- NEPS Scientific Use data is available free of charge!  
(after signing a data use agreement)