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
Using Stata for Disseminating NEPS SUF data

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
- nepsmiss

Installation:

```bash
net install nepstools, from("http://nocrypt.neps-data.de/stata")
```
3.1 \texttt{infoquery}:

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

- wrapper for \texttt{char list}
- presents predefined characteristics of variables
- syntax:
  \begin{verbatim}
  infoquery varlist [, alllanguages charlist(charlist)]
  \end{verbatim}
Using Stata for Disseminating NEPS SUF data
3 Accessing enriched metadata in Stata datasets

3.1 `infoquery`

**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[instancetype]:
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)]
3 Accessing enriched metadata in Stata datasets

3.2 `charren`

`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

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
```

3.3 \texttt{-nepsmiss-}

\textbf{nepsmiss: Motivation}

\begin{verbatim}
. 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)

\end{verbatim}

\begin{verbatim}
t514001_w1 -- Satisfied with life

\begin{tabular}{l|rrrr}
\hline
Valid & Freq. & Percent & Valid & Cum. \\
\hline
-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 & \\
\hline
\end{tabular}
\end{verbatim}
**nepsmis**: Overview

**nepsmis**: 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:
  ```
  nepsmis varlist [, reverse keeplabels misslist(numlist) verbose]
  ```
### Example

.nepsmiss t514001_w1

Warning: value label de935 is also attached to variables t514002_w1 [...]; > these surplus variables will not be recoded!
Consider running nepsmiss on those variables as well to correctly recode them!
Recoded 1979 values in total

.fre t514001_w1, tabulate(1)

<table>
<thead>
<tr>
<th></th>
<th>Freq.</th>
<th>Percent</th>
<th>Valid</th>
<th>Cum.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Valid 0 entirely dissatisfied</td>
<td>83</td>
<td>0.51</td>
<td>0.58</td>
<td>0.58</td>
</tr>
<tr>
<td>:</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10 entirely satisfied</td>
<td>2162</td>
<td>13.30</td>
<td>15.15</td>
<td>100.00</td>
</tr>
<tr>
<td>Total</td>
<td>14275</td>
<td>87.82</td>
<td>100.00</td>
<td></td>
</tr>
<tr>
<td>Missing .e Implausible value</td>
<td>20</td>
<td>0.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.j Unspecific missing</td>
<td>195</td>
<td>1.20</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.k Not participated</td>
<td>677</td>
<td>4.17</td>
<td></td>
<td></td>
</tr>
<tr>
<td>.m Missing by design</td>
<td>1087</td>
<td>6.69</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>1979</td>
<td>12.18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>16254</td>
<td>100.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
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 char files.
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
- NEPS Scientific Use data is available free of charge! (after signing a data use agreement)