Data and Metadata Management
at the German Socio-Economic Panel

Marcel Hebing (DIW Berlin)
marcel@ddionrails.org

Toronto, June 2014
Agenda

SOEP

Data

Metadata

Outlook
The **German Socio-Economic Panel (SOEP)** is a wide-ranging representative longitudinal study of private households, located at the German Institute for Economic Research, DIW Berlin. Every year, there were nearly **11,000 households**, and more than 20,000 persons sampled by the fieldwork organization TNS Infratest Sozialforschung.

The data provide information on all household members, consisting of Germans living in the Old and New German States, Foreigners, and recent Immigrants to Germany. The Panel was **started in 1984**.

Some of the many topics include household composition, occupational biographies, employment, earnings, health and satisfaction indicators.

http://www.diw.de/soep
Data
Data

- panel data (incl. spell data) on individual and household level
- wide format “SOEP-Core”
- new: long format “SOEPlong”
- innovation sample “SOEP-IS”
- related studies
Data access

- DVD, secure data download
- SOEPremote (via email, Lissy)
- remote access (first prototype)
- on site

- metadata portal “SOEPinfo” (basket, script generator)
File formats

- Stata
- SPSS
- SAS (switch to R)
- ASCII

- database: SIR

- long-term preservation: ASCII dump from SIR, incl. labels
- same dump is used for generation of Stata, SPSS, and SAS files.
Technical infrastructure

- database and processing: UNIX (AIX)
- metadata: Linux
- researchers and data managers: Windows/Linux

- Linux and UNIX: scripting!
Metadata
### 8.1.2 Party identification

#### General Party Identification

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1998</td>
<td>PP111</td>
<td>2008</td>
<td>2009</td>
<td>2010</td>
</tr>
<tr>
<td>2006</td>
<td>WP119</td>
<td>2011</td>
<td>BBP129</td>
<td>BAP128</td>
</tr>
</tbody>
</table>

© SOEP 2012
SOEPinfo v.2 will be based on the open source project DDI on Rails.
Vision. The data portal DDI on Rails accompanies researchers throughout the entire course of their research projects from conception to publication/citation.

The system offers researchers the possibility to explore the SOEP data, to compile personalized datasets, and to publish results on the publication database.
DDI on Rails = DDI + Ruby on Rails

- **Standard**: DDI (Data Documentation Initiative)
- **Web-framework**: Ruby on Rails
DDI Standard

- DDI Codebook (2.x): Import
- DDI Lifecycle (3.x): “Framework”
Architecture

Datasets (incl. labels) → r2ddi → Structure (CSV)

DDI on Rails
Understanding Data

Qlib by TNS

Endnote (XML)
Outlook
metadata-driven development

metadata-driven data processing
Links

- **Public beta:** https://data.soep.de
- **Internal testserver:** http://ddionrails.soep.de (user: guest, password: panel)
- **Project site:** http://www.ddionrails.org

Related Tools

- **r2ddi:** https://github.com/mhebing/r2ddi
- **soepR:** https://github.com/mhebing/soepR
- **testdata:** https://github.com/mhebing/testdata

Contact

- marcel@ddionrails.org
Thank you!