Documenting Panel Data Using DDI

# Same Same But Different

Knut Wenzig
EDDI16, Cologne

doi:10.5281/zenodo.167223, License: CC BY-SA 4.0 (exceptions see last slide)

# Abstract

The key characteristics of panel studies include repeated measures for a more or less stable sample over time. The core challenge in documenting panel studies is the documentation of these repeated measures (usually questions) and the resulting variables because various reasons can require modifications of measures over time – resulting in comparable but not identical data structures.

The DDI standard provides not one but multiple options for the documentation of panel data. In this workshop various options will be presented and their feasibility for common use cases will be discussed. The German Socio-Economic Panel (SOEP) will provide the primary use case, but participants are also invited to introduce and discuss their own use cases.

The workshop starts with a short introduction of both panel studies and the DDI standard. Therefore, no previous knowledge of the DDI standard is required to participate in the workshop. The goal for the workshop is to gain a deeper understanding of possible documentation strategies for panel studies.

DIW SOEP

Content

1. Introduction
2.
3.
4.
5.
6.
7.

# 1 Introduction

- What is today's topic?

- Participants
- Presenter

- What are the specific challenges of documenting a panel study?
- Participants: What are your challenges?

## I Am ...

- Name
- Institution
- Background
- Do you actually work with metadata?
- What do you expect from the workshop, would like to learn or discuss?

Hopefully the prepared slides fit to some of your expectations. It is possible to leave this path at any time!

## Your Ideas and Comments

- Please comment slides – each typo less counts!
- http://bit.do/samesame

- Easy with Google Presentations App

# Where I Am From ...

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 Kantar Public (formerly: 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

# Challenges, Specific to Panel Studies for Producers and Users

- What do “same” and “different” mean?
- Manage replication properly: re-use, don’t duplicate
- Identify replication
- Understand repeated measures
- Find corresponding variables
- Design data sets appropriate

Advanced:

- Measures change over time
- Understand variables which are
  - generated
  - transformed or
  - harmonized.
- Drive (parts of) data management using metadata.
- Connect with fieldwork and design.

# Shape of Data Sets – Pros and Cons

## Driven by logic of instruments

- One dataset per instrument
- One dataset per wave

## Driven by logic of analysis

- One dataset per wave
- Wide format (multiple waves)
- Long Format (multiple waves)

## Driven by logic of information

- One row per person
  - with all known persons
  - only with respondents
- One row per person and year
- One row per household, spell

Generally

- Which variables identifiy rows?
- Which universe?

## Type: Data Release ...

- … contains information from all waves
- or contains only information from one wave.
- Sometimes the term *version* is used with some sort of link to a wave or the number of waves.

## Same Same but Different. What Does This Mean?

- A movie
- Asian-English phrase
- Often a challenge, to decide whether two things are
  - the same: it's OK to substitute them mutually or use references
  - different: the information payload differs substantially
- In panel studies you want often to achieve the same but realize something different.

Content

1. Introduction
2. DDI Basics
3. Linking Information
4. Re-Use Information
5. Use Case: Framework@SOEP
6. Use Case: Data Management with Active Metadata
7. Use Case: Questionnaire Documentation

# Versions of DDI link

## DDI Codebook Latest Version: 2.5

- Scope: social science data documentation
- Recommended elements: DDI Lite (corresponds to 2.0)

## DDI Lifecycle Latest Version: 3.2

- Scope: data life cycle approach, social science data

## DDI Views

(Version 4, draft under review)

- Model-driven
- Functional views with subset of classes

# DDI Views (Version 4): Model-Driven Approach

- The model contains a *library* and *functional views* (subset for a specific purpose).
- The *library* is composed of *library packages* which contain other data types (primitives or complex) or classes.
- The *functional views* contain references to the classes used by the particular *functional view* that are needed to meet the needs of the use case or business application.
- The *functional views* loosely correspond to a DDI lifecycle business area.

Content

1. Introduction
2. DDI Basics
3. Linking Information
4. Re-Use Information
5. Use Case: Framework@SOEP
6. Use Case: Data Management with Active Metadata
7. Use Case: Questionnaire Documentation

# Linking Objects: Design Options in DDI Lifecycle (3.2)

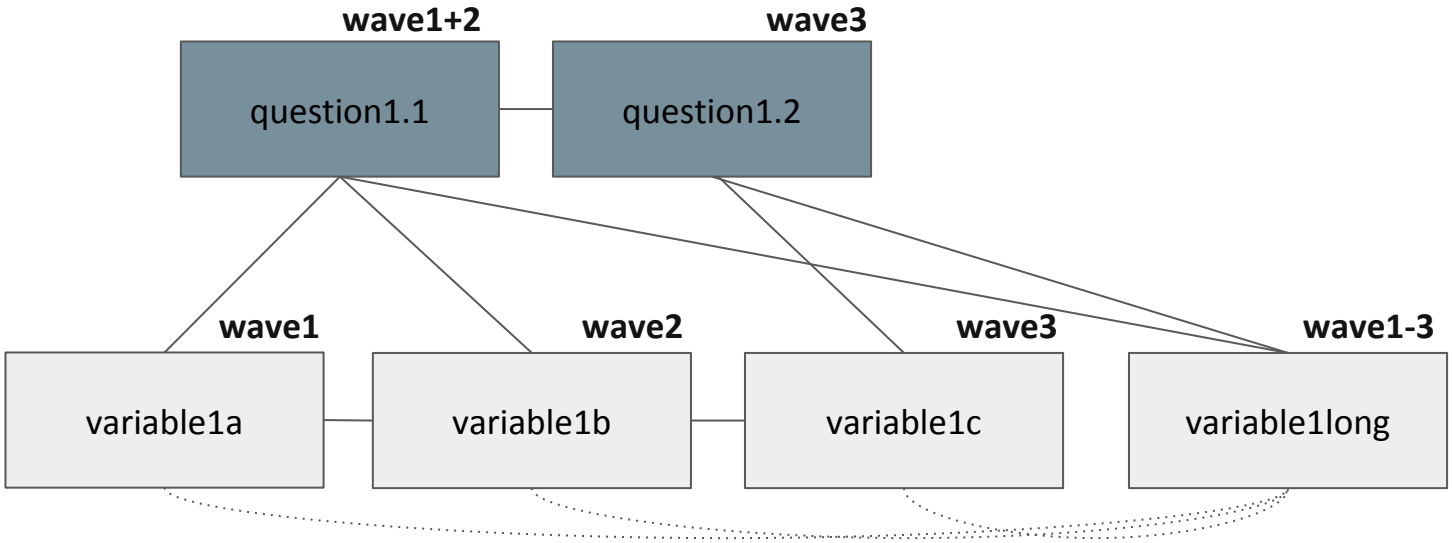
Direct Links

Versioning

Groups

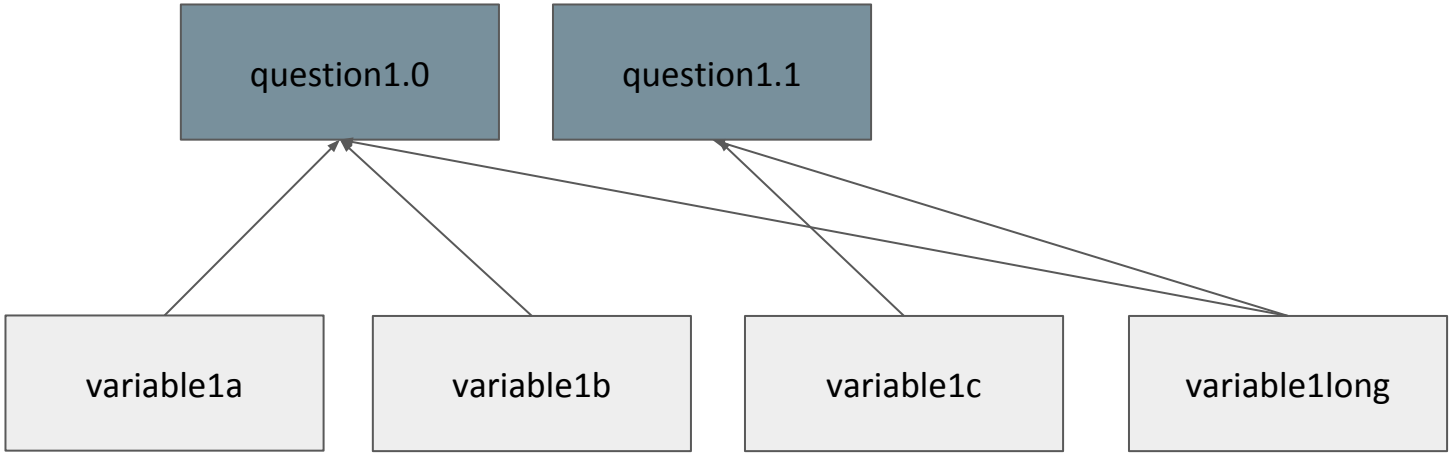
Concepts

# Some of the Hypothetical Connections




Only a selection – more with harmonized/derived variables.

# Link Variables to Questions Explicitly and Directly




- Links are specified from variable to questions

# A Variable Can Refer to One or More Questions

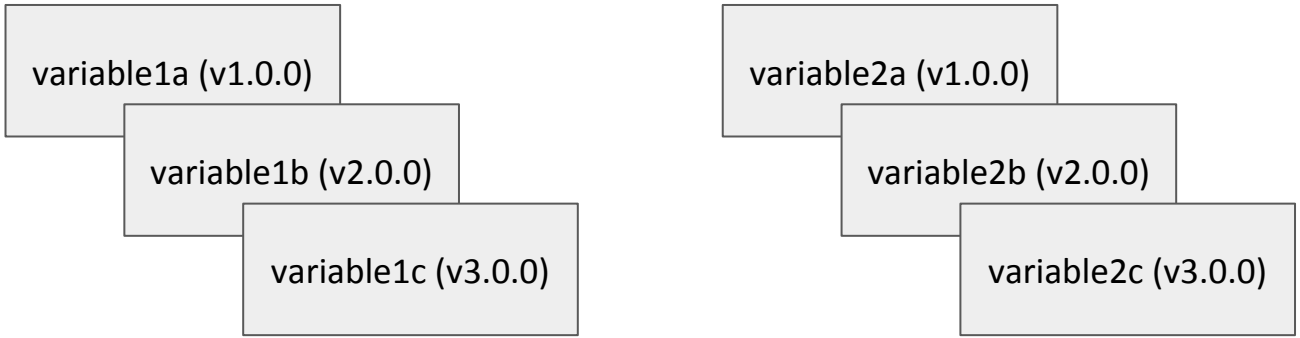
```
<l:Variable ...>
     <r:ID>variable1long</r:ID>
     <r:QuestionReference>
          <r:ID>question1.0</r:ID>
     </r:QuestionReference>
     <r:QuestionReference>
          <r:ID>question1.1</r:ID>
     </r:QuestionReference>

     ...
</l:Variable>
```

- Also possible, if a dataset results from different modes

- Links to the `field level documentation` in code examples

# Declare an Object as a New Version of Another Object




- Each panel wave corresponds to a version.
- Possible versioning rule: Wave.Minor.Sub-Minor
- BUT: versioning manages metadata object change over time

# A Variable Is Versionable

```
<l:Variable isVersionable="true" scopeOfUniqueness="Agency"
   versionDate="2012-10-31">
     <r:URN typeOfIdentifier="Canonical">
          urn:ddi:myagency.mypanel:variable1:2.0.0
     </r:URN>
     <VariableName>variable1b</VariableName>
     ...
</l:Variable>
```

- If variable names change over wave, ID will be an issue.
- `mypanel` is a sub-agency with agency `myagency`.
- Meaning in long datasets: “include all up to this wave”.
- Also possible for questions.

# Some Notes on Versioning in DDI

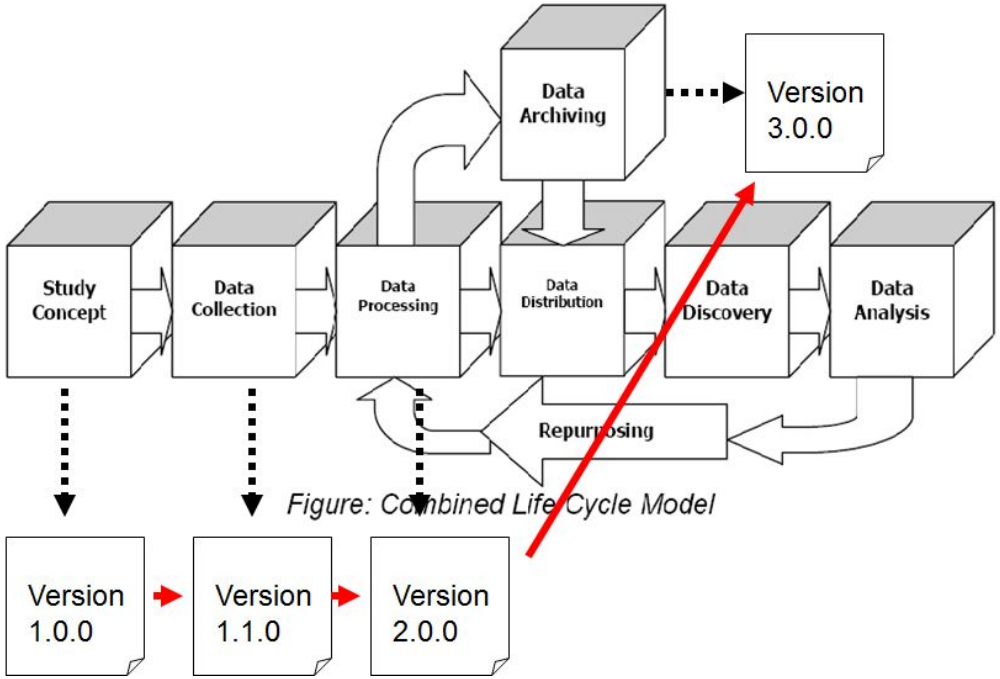
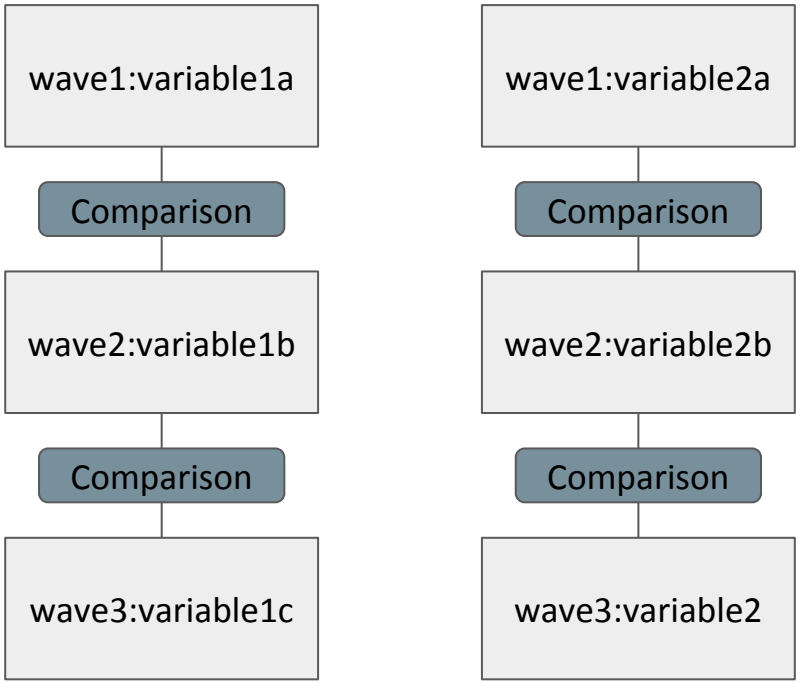



Figure: Combined Life Cycle Model

source: https://ddi-alliance.atlassian.net/wiki/download/attachments/491573/Just%20Enough%20DDI%203_Longitudinal.ppt

- No change after publication (new version)
- Elements: r:VersionRationale and r:BasedOnObject
- Business/technical versioning
- DDI Working Paper Series, Best Practises 8: Versioning and Publication, doi:10.3886/DDIBestPractices08

# Comparison Describes Relation of Two Objects

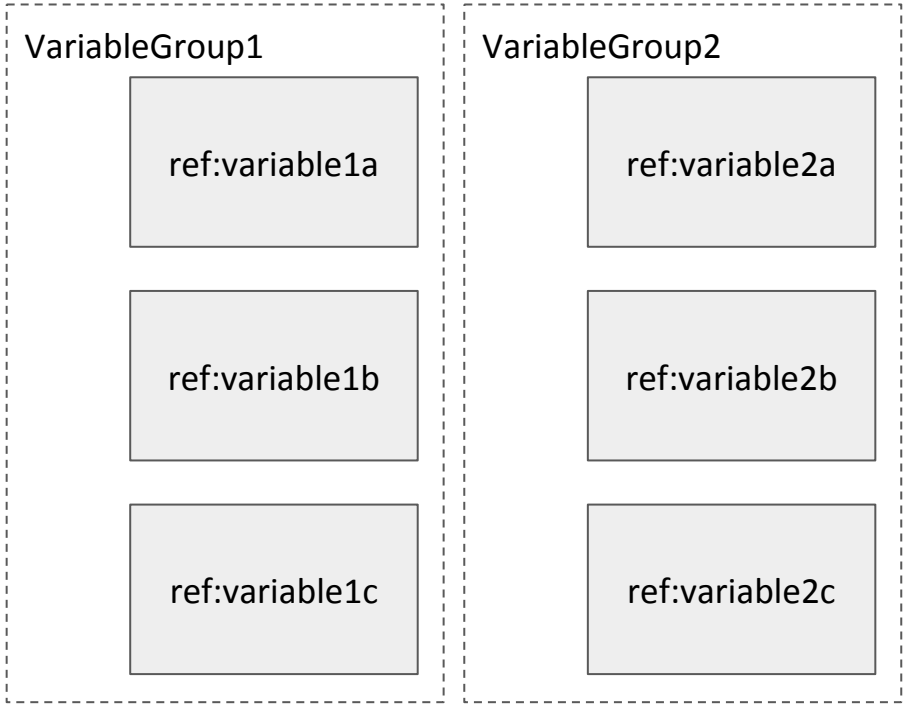



- Specified use: pairwise
- Many comparisons needed
- No re-use for comparisons
- Big correspondence table can be generated

# Comparison Can Hold Pairwise Maps

```
<c:VariableMap ...>
     <c:SourceSchemeReference>
          <r:ID>wave1:variable1a</r:ID><r:Version>2.0.0</r:Version>...
     </c:SourceSchemeReference>
     <c:TargetSchemeReference>
          <r:ID>wave2:variable1b</r:ID><r:Version>2.0.0</r:Version>...
     </c:TargetSchemeReference>
     <c:Correspondence>
          <c:Commonality ...>
               Target (wave2) is a repeated measure of source (wave1)
          </c:Commonality>
          <c:Difference ...>...</c:Difference>
     </c:Correspondence>
     ...
</c:VariableMap>
<c:QuestionMap ...>...</c:QuestionMap>
```

# Tie VariableReferences Together with a VariableGroup




- Virtual, only references are used
- Using Group is also possible

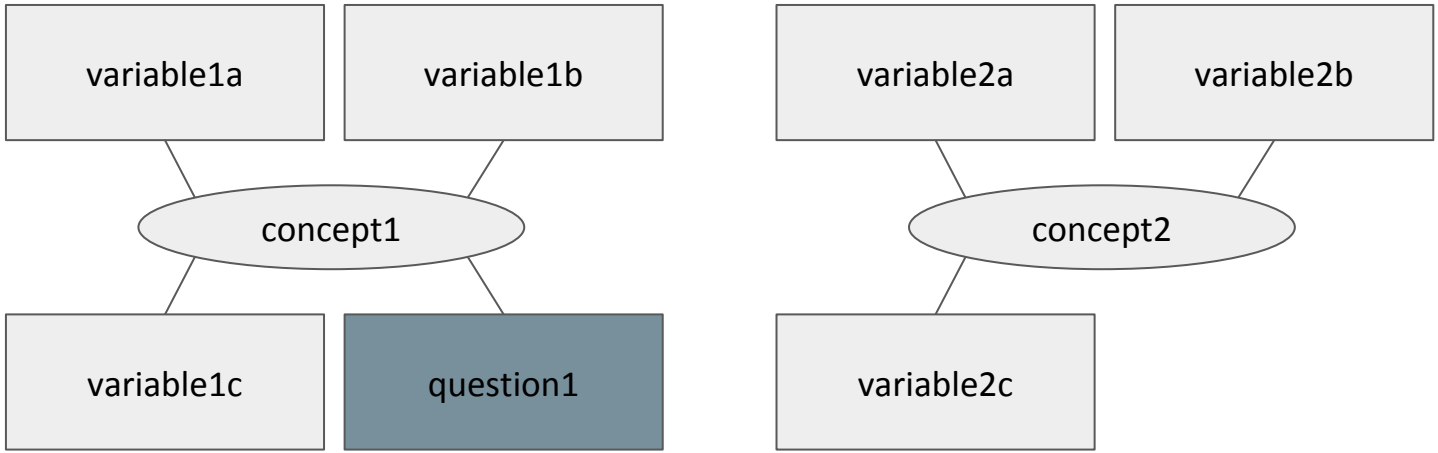
(Example showing the grouping approach for comparable variables, description of derived variables as well as the relationship of waves and the household/person relationship)

# VariableGroup Is a List of References

```
<l:VariableGroup ...>
     <r:VariableReference>
          <r:ID>wave1:variable1a</r:ID>
          <r:Version>3.0.0</r:Version>...
     </r:VariableReference>
     <r:VariableReference>
          <r:ID>wave2:variable1b</r:ID>
          <r:Version>3.0.0</r:Version>...
     </r:VariableReference>
     <r:VariableReference>
          <r:ID>wave3:variable1c</r:ID>
          <r:Version>3.0.0</r:Version>...
     </r:VariableReference>
     ...
</l:VariableGroup>
```

- Can have a name
- QuestionGroup also available

# Tag an Object with a Concept




- Works for variables and questions at the same time

# ConceptReference Used in Variable and QuestionItem

```
<l:Variable ...>
     <r:ID>wave1:variable1a</r:ID>
     <r:Version>3.0.0</r:Version>
     <r:ConceptReference>
          <r:ID>myconcept1</r:ID>
          ...
     </r:ConceptReference>
     ...
</l:Variable>
```

```
<d:QuestionItem ...>
     <r:ID>wave1:question1</r:ID>
     <r:Version>1.0.0</r:Version>
     <r:ConceptReference>
          <r:ID>myconcept1</r:ID>
          ...
     </r:ConceptReference>
     ...
</l:QuestionItem>
```

## Which Method Would You (Not) Use, If ...

- Messurement is stable or changes a lot over time?
- Information is stored in a long format, i.e. one variable contains information from multiple waves?
- Transformation information should be stored?
- Versioning of metadata differs from data?

- Your use case?

## Choosing the Right Method

- depends on ressources you have for the work to be done,
- on amount and structure of information you want to store.
- Direction and integration of references differ!
- Proof of concept will be necessary.
- Take mix of methods into account.

The workflow in your organisation will change!

Content

1. Introduction
2. DDI Basics
3. Linking Information
4. Re-Use Information
5. Use Case: Framework@SOEP
6. Use Case: Data Management with Active Metadata
7. Use Case: Questionnaire Documentation

## What do we want to achieve?

- Especially in panel studies:
  Repetition has to be managed. (Only if you use a reference, you know, that you use exactly the same.)
- Something used more than once?
  Try to re-use by reference, do not copy.
- Great amount of questions and variables: great amount of metadata to manage.

# Answers within a Question refer to a CodeList

```
<d:QuestionItem ...>
      ...
      <d:CodeDomain>
            <r:CodeListReference>
                <r:ID>codeList1</r:ID>
            </r:CodeSchemeReference>
      </d:CodeDomain>
</d:QuestionItem>
```

- Already usefull for one single study
- Various other possibilities for ResponseDomain
- CodeList can contain CodeListReferences

# Value Labels of Variables Refer to a CodeList

```
<l:Variable ...>
     ...
     <l:CodeRepresentation>
          <r:CodeListReference>
               <r:ID>codeList2</r:ID>
          </r:CodeListReference>
     </l:CodeRepresentation>
</l:Variable>
```

- Already usefull for one single study
- CodeRepresentation is contained by substitution in VariableRepresentation
- CodeList can contain CodeListReferences

## CodeLists for Questions and Variables

Q5: Are you happy today?

[1] Yes

[2] No

[-1] Don’t know

- -2 due to routing
- -8 long dataset

VAR5: Happiness day of int.

[1] Yes

[2] No

[-1] Don’t know

[-2] does not apply/not reached

[-8] missing due to design

# CodeLists for Questions and Variables

## Q5: Are you happy today?

CodeList1
[1] Yes
[2] No

CodeList2
[-1] Don’t know

CodeListQ5

## VAR5: Happiness day of int.

CodeList1
[1] Yes
[2] No

CodeList3
[-1] Don’t know
[-2] does not apply/not reached
[-8] missing due to design

CodeListVAR5

# CodeLists for Questions and Variables

CodeList1
[1] Yes
[2] No

CodeList2
[-1] Don’t know

CodeList3
[-1] Don’t know
[-2] does not apply/not reached
[-8] missing due to design

CodeList1: substantial values (used in questions and variables)

CodeList2: missing values (extension used in questions)

CodeList3: missing values (used in data, reuseable)

# Questions: Multiple Levels of Reference and Re-Use

- < DataCollection
  - < QuestionSchemeReference
- < QuestionScheme
  - < QuestionSchemeReference
  - < QuestionGroupReference
  - < QuestionBlockReference
  - < QuestionGridReference
  - < QuestionItemReference
- < QuestionGroup
  - < QuestionGroupReference
  - < QuestionBlockReference
  - < QuestionGridReference
  - < QuestionItemReference
- < QuestionBlock
  - < QuestionItemReference

## Partial Re-Use of Main Elements

- Variable
  - Name
  - Label
  - Value Labels/ CodeList
  - Concept
- Question
  - Text
  - Answers/Code List
  - Concept
- If variable names (or question text) change: Can schemes combined with inheritance could be a solution?

## There is one disadvantage, too:

- If you change an object, which is refered to, you risk to change each object, that uses it.
- This may not always be what you want.
- Check where the object is used before you change it.

## And Always One Conflict of Objective:

- The more information adressed by one single reference,
  - the less information you have to state additionally (good)
  - the more often you have to specifiy new objects (bad).

Content

1. Introduction
2. DDI Basics
3. Linking Information
4. Re-Use Information
5. Use Case: Framework@SOEP
6. Use Case: Data Management with Active Metadata
7. Use Case: Questionnaire Documentation

# Use Case: Framework@SOEP

## paneldata.org (driven by DDIonRails)

- Successor for SOEPinfo, needed due to design of file structure (one file per wave with changing variable names).
- Possibility to explore the data, and to compile personalized datasets.
- Multiple Studies (hosted service for other panel studies)
- Linking across Studies (using concepts)
- Panel-specific functionality

## paneldata.org (and Other Tools) Use Standardized Tables

- paneldata.org driven by DDIonRails 1/2 which can be understood as an implemetation of DDI.
- paneldata.org a search tool for metadata of panel surveys (no metadata curation on this platform)
- DDI's XML structure extremely simplified and flattened to relational tables, which preserve selected features.
- Input: Markdown Files, Datasets (Stata) and Tables

## Tables with Metadata for paneldata.org

- Contain information on instruments, connections (questions/variables, variables/variables) and concepts of variables
- Tables are stored in CSV files, which turned out to be easily editable by students and apprentices.
- Git version control helps a lot (collaboration!)
- Displayed online (paneldata.org) and in PDFs (R/LaTeX)
- Extremely economic set-up

## DDI Structures We Use

- Link from variables (in raw data) to questions like DDI QuestionReference in DDI Variable
- Link pairs of variables (DDI VariableMap)
  - raw to published data
  - raw to generated data
  - consolidate data from two or more questionnaires (mode, long)
- Identify repeated measurement (DDI Concept)
  - Questions and Variables (item correspondence)
- Re-Use answers within a questionnaire

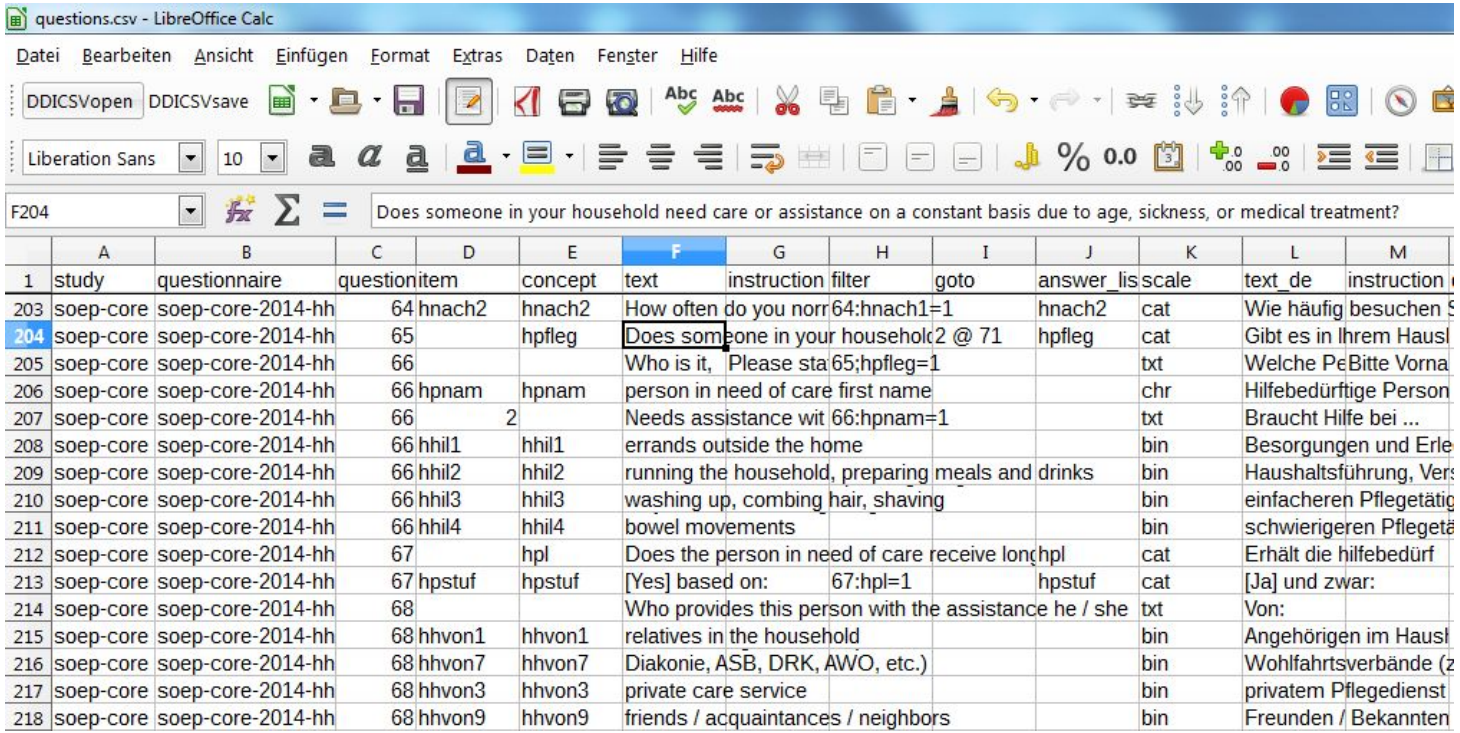
# Use Case: Framework@SOEP

## Just Introduced (outside paneldata.org)

- System-wide use of CodeLists for questions and variables (valid and missing values)
  - order
  - Value (Code)
  - Label (Category)
- System-wide use of templates (scheme) for questions
  - QuestionText
  - InterviewerInstruction
  - ResponseDomain
  - Concept

Helps to reduce cost of translation.

# Use Case: Framework@SOEP

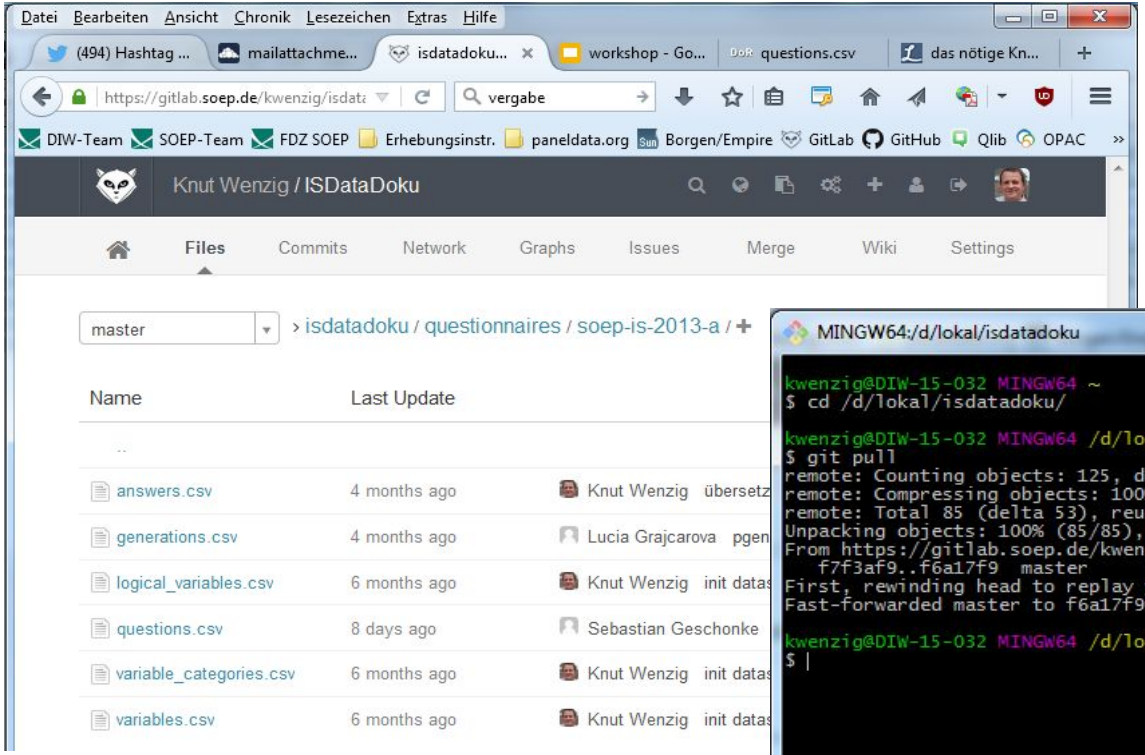


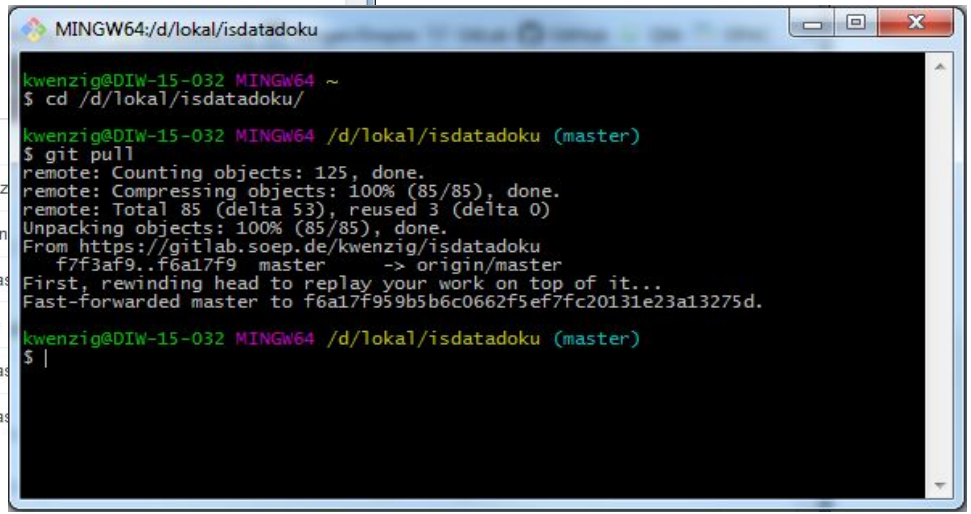

| | A | B | C | D | E | F | G | H | I | J | K | L | M |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | study | questionnaire | question | item | concept | text | instruction | filter | goto | answer_lis | scale | text_de | instruction |
| 203 | soep-core | soep-core-2014-hh | 64 | hnach2 | hnach2 | How often do you norr | | 64:hnach1=1 | | hnach2 | cat | Wie häufig | besuchen S |
| 204 | soep-core | soep-core-2014-hh | 65 | | hpfleg | Does someone in your household | | | 2 @ 71 | hpfleg | cat | Gibt es in Ihrem Hausl | |
| 205 | soep-core | soep-core-2014-hh | 66 | | | Who is it, | Please sta | 65;hpfleg=1 | | | txt | Welche Pe | Bitte Vorna |
| 206 | soep-core | soep-core-2014-hh | 66 | hpnam | hpnam | person in need of care | first name | | | | chr | Hilfebedürftige Person | |
| 207 | soep-core | soep-core-2014-hh | 66 | 2 | | Needs assistance wit | | 66:hpnam=1 | | | txt | Braucht Hilfe bei ... | |
| 208 | soep-core | soep-core-2014-hh | 66 | hhil1 | hhil1 | errands outside the home | | | | | bin | Besorgungen und Erle | |
| 209 | soep-core | soep-core-2014-hh | 66 | hhil2 | hhil2 | running the household, preparing meals and | | | | drinks | bin | Haushaltsführung, Vers | |
| 210 | soep-core | soep-core-2014-hh | 66 | hhil3 | hhil3 | washing up, combing hair, shaving | | | | | bin | einfacheren Pflegetätig | |
| 211 | soep-core | soep-core-2014-hh | 66 | hhil4 | hhil4 | bowel movements | | | | | bin | schwierigeren Pflegetä | |
| 212 | soep-core | soep-core-2014-hh | 67 | | hpl | Does the person in need of care receive lon | | | | hpl | cat | Erhält die hilfebedürf | |
| 213 | soep-core | soep-core-2014-hh | 67 | hpstuf | hpstuf | [Yes] based on: | | 67:hpl=1 | | hpstuf | cat | [Ja] und zwar: | |
| 214 | soep-core | soep-core-2014-hh | 68 | | | Who provides this person with the assistance | | | | he / she | txt | Von: | |
| 215 | soep-core | soep-core-2014-hh | 68 | hhvon1 | hhvon1 | relatives in the household | | | | | bin | Angehörigen im Hausl | |
| 216 | soep-core | soep-core-2014-hh | 68 | hhvon7 | hhvon7 | Diakonie, ASB, DRK, AWO, etc.) | | | | | bin | Wohlfahrtsverbände (z | |
| 217 | soep-core | soep-core-2014-hh | 68 | hhvon3 | hhvon3 | private care service | | | | | bin | privatem Pflegedienst | |
| 218 | soep-core | soep-core-2014-hh | 68 | hhvon9 | hhvon9 | friends / acquaintances / neighbors | | | | | bin | Freunden / Bekannten | |

**Screenshot: Questionnaire in LibO Calc with two new buttons**

DIW SOEP

# Use Case: Framework@SOEP




## Gitlab and Git Bash

# CSV files on Git

## Pro

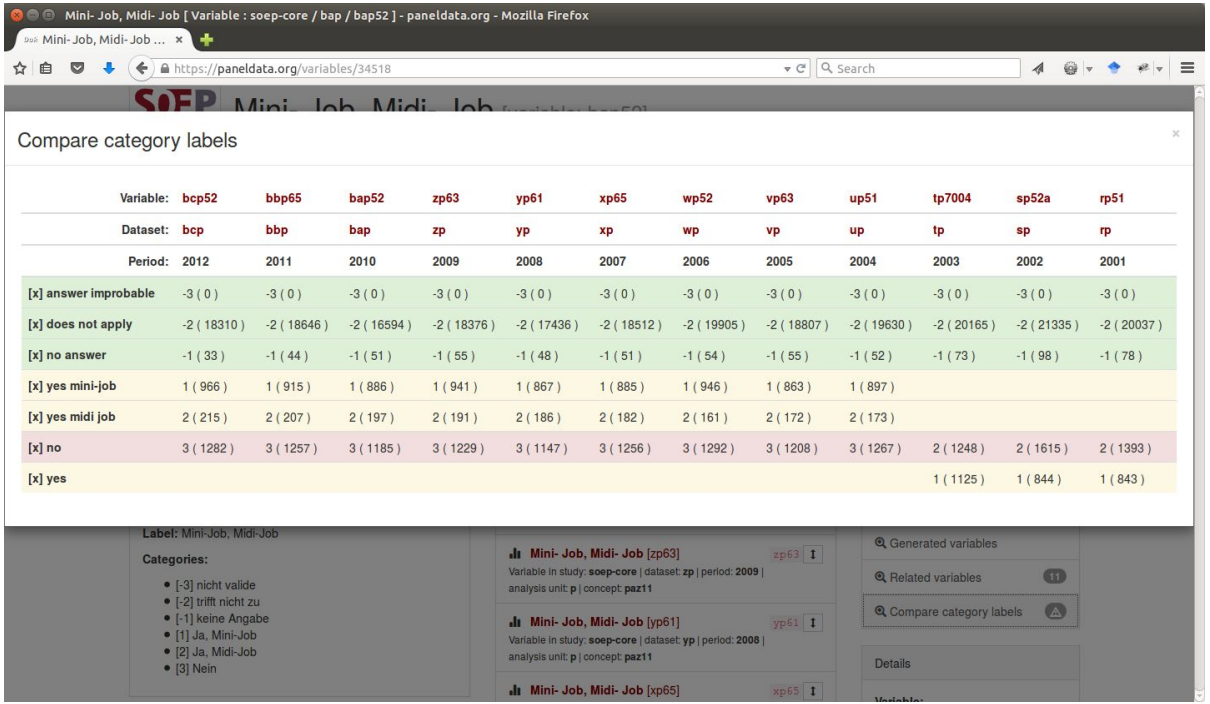
- No server only software on clients needed (but Gitlab or similar make things easier)
- No special frontend (“editor”) needed: lean development
- Version control helps to track changes and reset to previous version in case of errors
- Metadata easy accessible for programming (Ruby, R, Stata)
- Establish version control know-how

## Con

- Integrity of metadata not enforced
- Annoying issues with seperators, encoding, quotes (LibO Calc and a macro helps)
- Transfer to database (for web-use)No server only software on clients needed (but Gitlab or similar make things easier)

Do you already use version control?

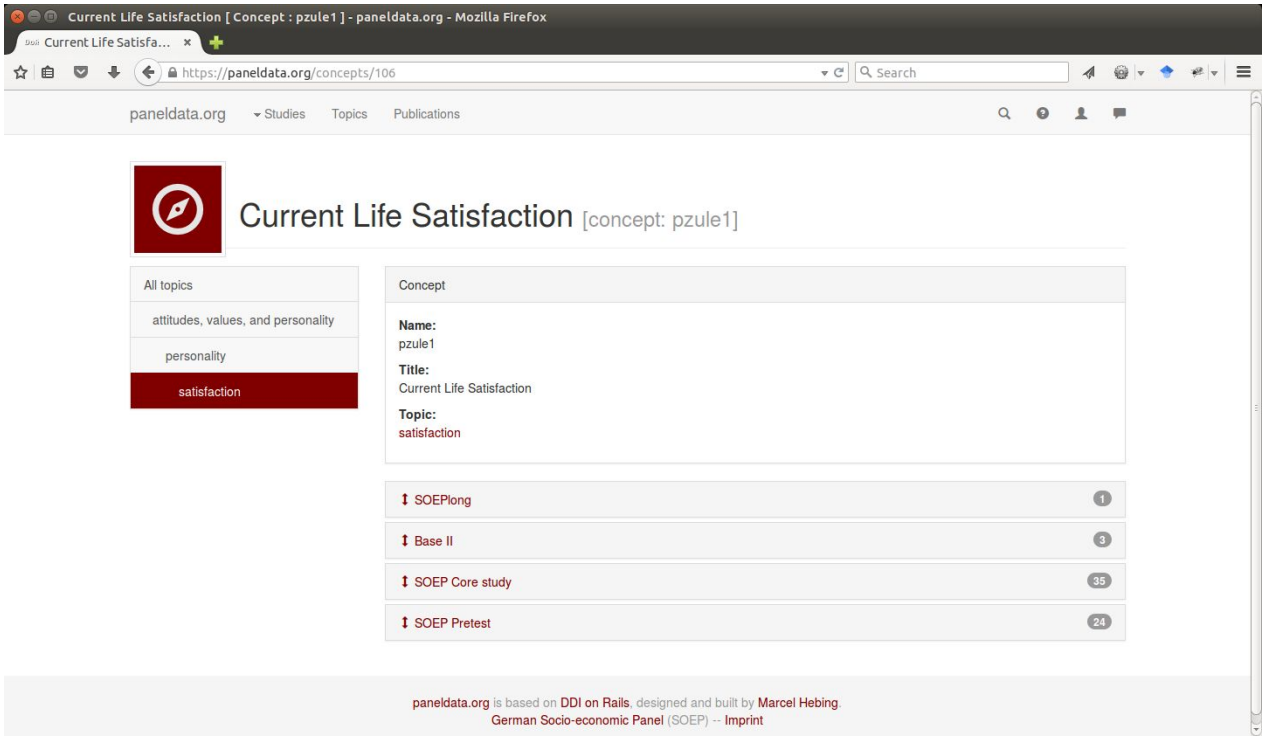
# Use Case: Framework@SOEP




| Variable: | bcp52 | bbp65 | bap52 | zp63 | yp61 | xp65 | wp52 | vp63 | up51 | tp7004 | sp52a | rp51 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Dataset: | bcp | bbp | bap | zp | yp | xp | wp | vp | up | tp | sp | rp |
| Period: | 2012 | 2011 | 2010 | 2009 | 2008 | 2007 | 2006 | 2005 | 2004 | 2003 | 2002 | 2001 |
| [x] answer improbable | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) | -3 ( 0 ) |
| [x] does not apply | -2 ( 18310 ) | -2 ( 18646 ) | -2 ( 16594 ) | -2 ( 18376 ) | -2 ( 17436 ) | -2 ( 18512 ) | -2 ( 19905 ) | -2 ( 18807 ) | -2 ( 19630 ) | -2 ( 20165 ) | -2 ( 21335 ) | -2 ( 20037 ) |
| [x] no answer | -1 ( 33 ) | -1 ( 44 ) | -1 ( 51 ) | -1 ( 55 ) | -1 ( 48 ) | -1 ( 51 ) | -1 ( 54 ) | -1 ( 55 ) | -1 ( 52 ) | -1 ( 73 ) | -1 ( 98 ) | -1 ( 78 ) |
| [x] yes mini-job | 1 ( 966 ) | 1 ( 915 ) | 1 ( 886 ) | 1 ( 941 ) | 1 ( 867 ) | 1 ( 885 ) | 1 ( 946 ) | 1 ( 863 ) | 1 ( 897 ) | | | |
| [x] yes midi job | 2 ( 215 ) | 2 ( 207 ) | 2 ( 197 ) | 2 ( 191 ) | 2 ( 186 ) | 2 ( 182 ) | 2 ( 161 ) | 2 ( 172 ) | 2 ( 173 ) | | | |
| [x] no | 3 ( 1282 ) | 3 ( 1257 ) | 3 ( 1185 ) | 3 ( 1229 ) | 3 ( 1147 ) | 3 ( 1256 ) | 3 ( 1292 ) | 3 ( 1208 ) | 3 ( 1267 ) | 2 ( 1248 ) | 2 ( 1615 ) | 2 ( 1393 ) |
| [x] yes | | | | | | | | | | 1 ( 1125 ) | 1 ( 844 ) | 1 ( 843 ) |

## Change of Categories over Time (DDI Concept)

# Use Case: Framework@SOEP




## Linking across Studies (using DDI Concept)

Content

1. Introduction
2. DDI Basics
3. Linking Information
4. Re-Use Information
5. Use Case: Framework@SOEP
6. Use Case: Data Management with Active Metadata
7. Use Case: Questionnaire Documentation

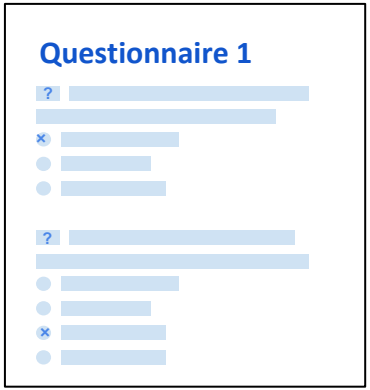
# Consolidate Information from Two Questionnaires

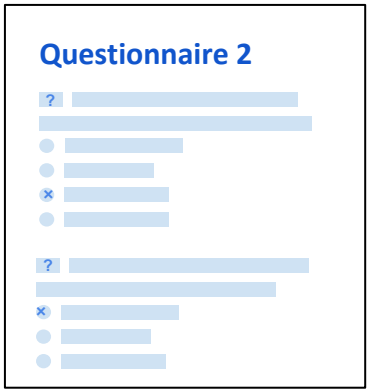
SOEP

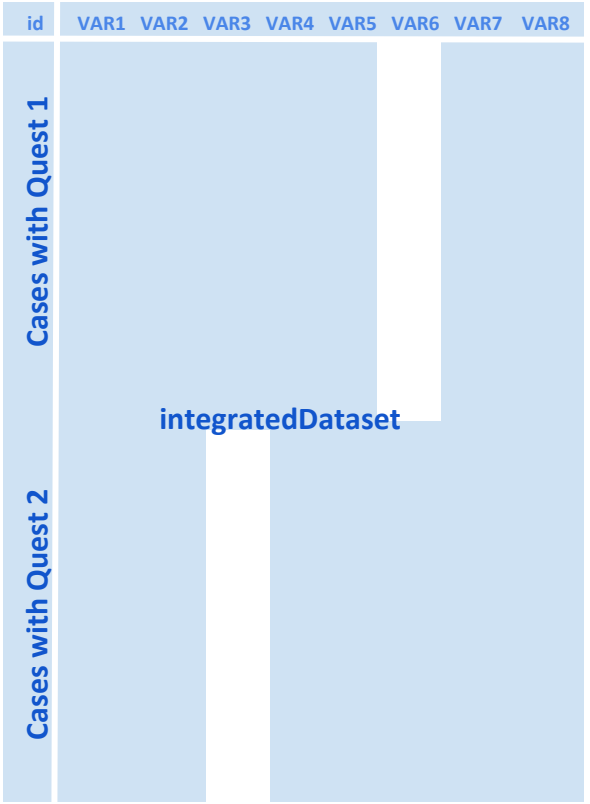
FiD

- Pool information of two very similar studies which were carried out in the 4 same years
  - SOEP
  - Families in Germany (FiD)
- Integration reduces burden of data users dramatically
  - identification of similar questions/variables
  - harmonisation of information is standardised
- Very similar use case: integration of datasets from different waves/modes

# Use Case: Data Management with Active Metadata





# Use Case: Data Management with Active Metadata

- Identify corresponding questions/variables
- Correct, (harmonise)
- Rename variables:
  - Dataset1, var1
    > integratedDatset, VAR1
  - See table
- Compare corresponding variables
  - Prevent errors
  - Variable labels
  - Value labels
  - Accept differences/make corrections
- Append datasets
  - Fill sparse areas with missing code
- Evaluate work
- (Harmonise)

| i_dataset | i_variable | o_dataset | o_variable |
|---|---|---|---|
| Dataset1 | var1 | iDataset | VAR1 |
| Dataset1 | var2 | iDataset | VAR2 |
| Dataset2 | var1 | iDataset | VAR1 |
| Dataset2 | var2 | iDataset | VAR2 |

**Table equivalent to many DDI VariableMaps**

# Use Case: Data Management with Active Metadata

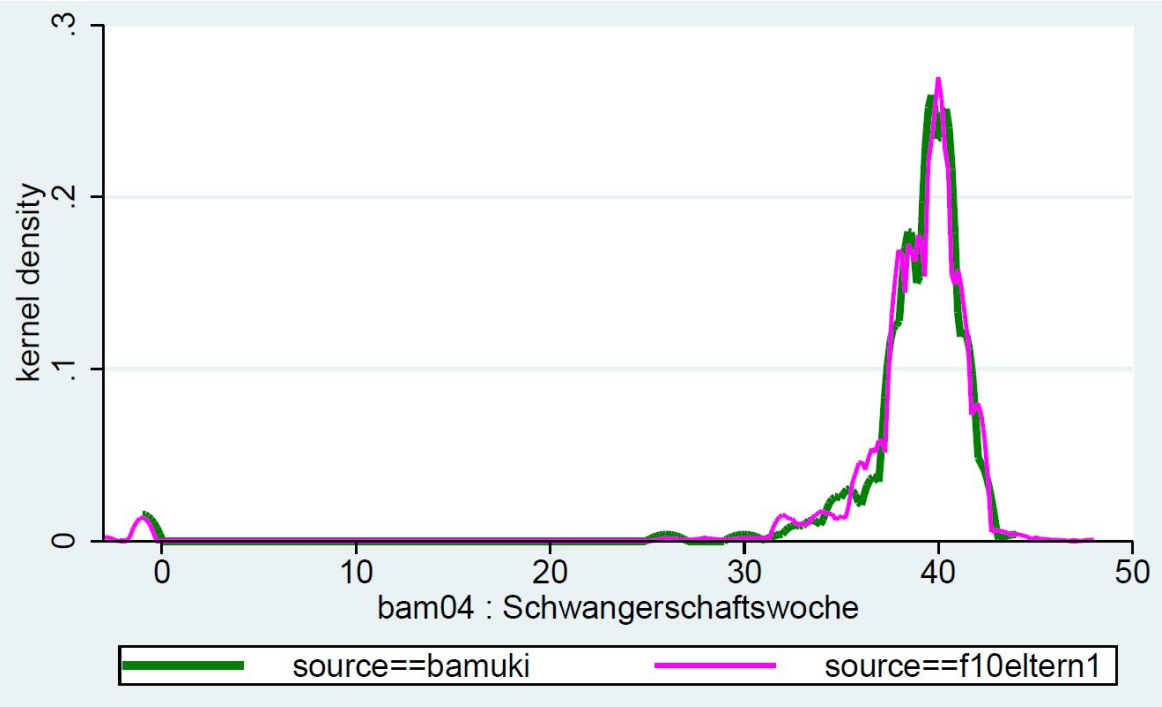
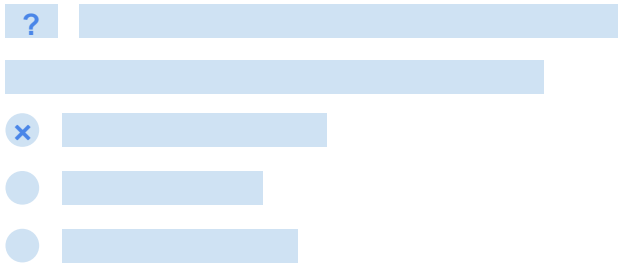



Diagramm shows successful integration of a metric variable (week of pregnancy)

# Use Case: Data Management with Active Metadata

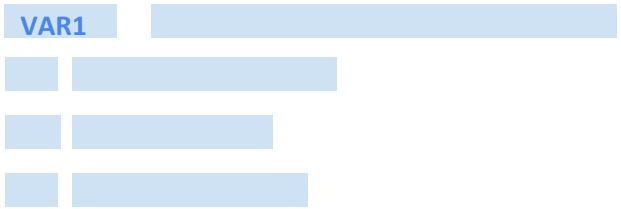
## Questionnaire 1




**See:**
**Dataset1, var1**
**integratedDatatset, VAR1**

original renaming information – nothing more

## Codebook: integratedDataset




**Sources:**
**Questionnaire 1, Q12**
**Questionnaire 2, Q14**

# Use Case: Data Management with Active Metadata

## Result

- 62 Stata files with integrated information
- 305 lines of code (without corrections)
- 21915 (non-)renaming of variables
- 61464 differences in variable labels and value labels were accepted
- Stata ados which rely on DDIonRails metadata: http://ddionrails.org/stata/
  - dorrename, dorcomparedta, dorcomparexls, dorappend, dorevaluate, dororder, dorlabeldta
  - https://github.com/ddionrails/stata

## Metadata driven processing

- Code written for data preparation more structured and with less lines and better to maintain
- Metadata (and documentation) more accurate
- Documentation ready when data are ready

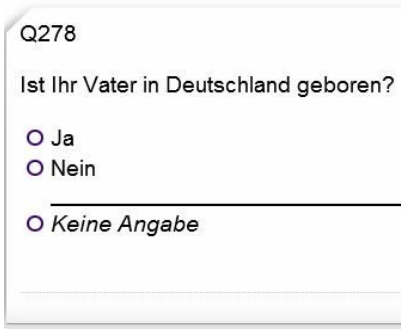
Content

1. Introduction
2. DDI Basics
3. Linking Information
4. Re-Use Information
5. Use Case: Framework@SOEP
6. Use Case: Data Management with Active Metadata
7. Use Case: Questionnaire Documentation

# Use Case: Questionnaire Documentation

## Questionnaire Metadata

- Reference material
- Some information is essential (has to be preserved, depends on use case)
- Some information can be ignored (not captured by metadata)
- i18n: multilingual infrastructure (fieldwork and/or documentation)
- Re-use of information (next wave)




<< CAPI Screen

PAPI

65. **Does someone in your household need care or assistance on a constant basis due to age, sickness, or medical treatment?**

Yes ☐ ⇩ No ☐ ➡ *Question 71!*

66. **Who is it, and which of the following activities does he or she need assistance in?**

☞ *Please state the person's first name. If there is more than one person in need of care in the household, please state the person most in need of care.*

**person in need of care first name** ☐

**Needs assistance with ...**

errands outside the home ☐

running the household, preparing meals and drinks ☐

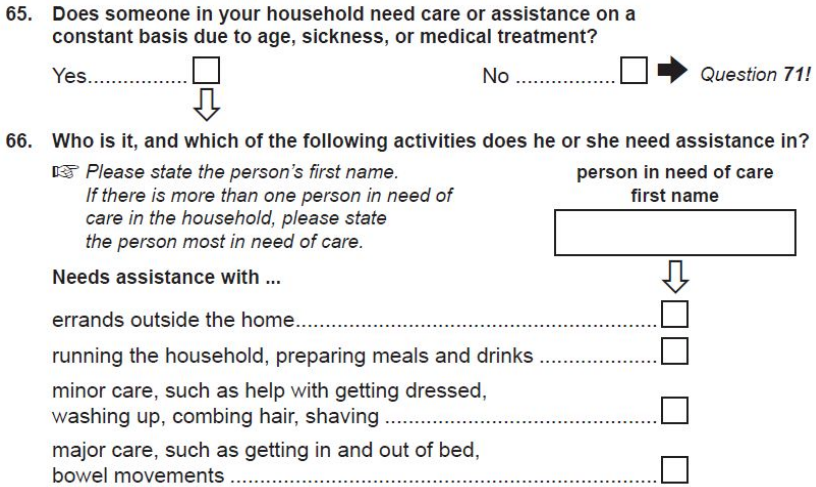
minor care, such as help with getting dressed, washing up, combing hair, shaving ☐

major care, such as getting in and out of bed, bowel movements ☐

# Use Case: Questionnaire Documentation

## (Not) preserved information

### Source Material (Paper)



65. **Does someone in your household need care or assistance on a constant basis due to age, sickness, or medical treatment?**

Yes ☐ ⇩ No ☐ ➡ *Question 71!*

66. **Who is it, and which of the following activities does he or she need assistance in?**

☞ *Please state the person's first name. If there is more than one person in need of care in the household, please state the person most in need of care.*

**person in need of care first name** ☐

**Needs assistance with ...**

errands outside the home ☐

running the household, preparing meals and drinks ☐

minor care, such as help with getting dressed, washing up, combing hair, shaving ☐

major care, such as getting in and out of bed, bowel movements ☐

### Produced with Metadata

**65 Does someone in your household need care or assistance on a constant basis due to age, sickness, or medical treatment?**

Yes 1
No 2

65;hpfleg hpfleg 2 @ 71

65;hpfleg=1

**66 Who is it, and which of the following activities does he or she need assistance in?**

*Please state the person's first name. If there is more than one person in need of care in the household, please state the person most in need of care.*

person in need of care first name

66:hpnam hpnam

Needs assistance with ...

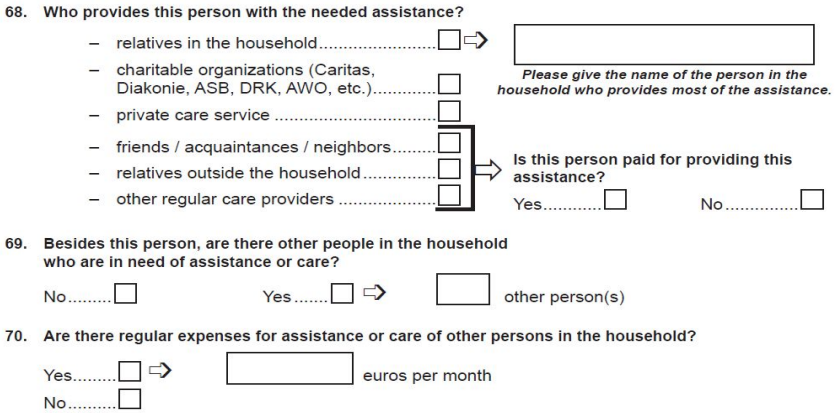
| | |
|---|---|
| errands outside the home | 1 |
| running the household, preparing meals and drinks | 1 |
| minor care, such as help with getting dressed, washing up, combing hair, shaving | 1 |
| major care, such as getting in and out of bed, bowel movements | 1 |

66:hhil1 hhil1
66:hhil2 hhil2
66:hhil3 hhil3
66:hhil4 hhil4

# Use Case: Questionnaire Documentation

## (Not) preserved information

### Source Material (Paper)



**68. Who provides this person with the needed assistance?**

- relatives in the household ☐ ⇨
- charitable organizations (Caritas, Diakonie, ASB, DRK, AWO, etc.) ☐
- private care service ☐
- friends / acquaintances / neighbors ☐
- relatives outside the household ☐
- other regular care providers ☐

*Please give the name of the person in the household who provides most of the assistance.*

**Is this person paid for providing this assistance?**

Yes ☐ No ☐

**69. Besides this person, are there other people in the household who are in need of assistance or care?**

No ☐ Yes ☐ ⇨ ☐ other person(s)

**70. Are there regular expenses for assistance or care of other persons in the household?**

Yes ☐ ⇨ ☐ euros per month

No ☐

Produced with Metadata >>

**68 Who provides this person with the assistance he / she needs?**

| | |
|---|---|
| relatives in the household | 1 |
| charitable organizations (Caritas, Diakonie, ASB, DRK, AWO, etc.) | 1 |
| private care service | 1 |
| friends / acquaintances / neighbors | 1 |
| relatives outside the household | 1 |
| other regular care providers | 1 |

68:hhvon1 hhvon1
68:hhvon7 hhvon7
68:hhvon3 hhvon3
68:hhvon9 hhvon9
68:hhvon6 hhvon6
68:hhvon8 hhvon8

[relatives in the household] Please give us the name of the person in the household who is the main caregiver.
68:hhnam hhnam 68:hhvon1=1

[friends / acquaintances / neighbors, relatives outside the household,other regular care providers]Is this person paid for providing this assistance?

Yes 1
No 2

68:hhbez hhbez 68:hhvon9=1 | 68:hhvon6=1 | 68:hhvon8=1

**69 Besides this person, are there other people in the household who are in need of assistance or care?**

No 2
Yes 1

69;hpfleg2 hpfleg2

[Yes] ... other person(s)
69:hpfleg3 hpfleg3 69;hpfleg2=1 hpfleg2=1

**70 Are there regular expenses for assistance or care of other persons in the household?**

Yes 1
No 2

70;hpfleg4 hpfleg4

Yes ... euros per month
70:hpfleg5 hpfleg5 70;hpfleg4=1

# Example: What DDIonRails preserves and adds

Preserved

- Question numbers
- Textual information (question texts, instructions, answers)
- Routing (logical: filter, goto)

Added

- Values for answers
- Concepts
- Links to variables (DDI QuestionsReference)
- Translations

Not preserved

- Layout (horizontal/vertical arrangement, text prior/after open ended questions)
- Typography (bold, underlined)
- Graphical information
- Routing (textual)

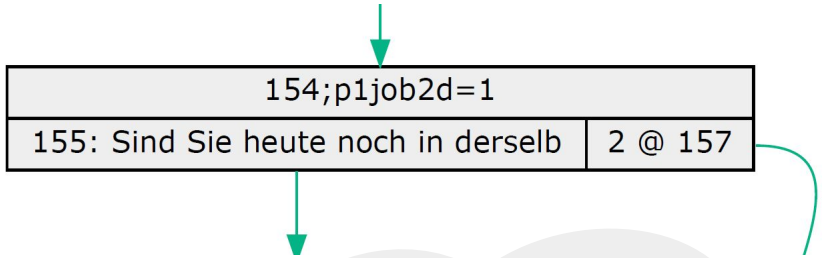
What information do you want to preserve?

# Some Notes on Routing

- Common default: go to next question
  - No more specification needed
  - Exceptions needed

Two different approaches in instruments:

- Question's gatekeeper ("filter")
  - Defines the universe of this particular question
  - Condition which has to be true
- After a question ("goto")
  - Defines the way to the next question depending on the answer (and perhaps other information)




- Which approach is used in your institution? What are your experiences?
- What do data users like, what survey designers – and why?
- Which approach is more, which is less parsimonious?
- What about visualization?
- Will it convert?

# Routing in DDI

ControlConstruct:

Extensible structure for control elements used in describing flow logic within the instrument: IfThenElse, RepeatUntil, RepeatWhile, Loop, Sequence, ComputationItem, StatementItem, and QuestionConstruct. (from DDI 3.2 XML Schema Documentation)

```
<d:IfThenElse>
    <d:IfCondition>
        <r:Code programmingLanguage="Neutral">Counter != 1</r:Code>
    </d:IfCondition>
    <d:ThenConstructReference>
        <r:ID>333ae135-784d-4435-9e54-...</r:ID>
    </d:ThenConstructReference>
</d:IfThenElse>
```

source: http://www.colectica.com/census2010-ddi-metadata (shortened, DDI 3.1)

Some kind of code needed to specify conditions and or calculations, to some amount defined within DDI.

- Operators
- References
  - OutParameter
  - InParameter
  - Binding
- Re-use difficult by nature
- Re-use on Fragment level

# Example: Routing in DDIonRails

- Each item (one item is related to one variable) in a question can have a filter and a goto.
- A filter can have references to one or more (prior) items in the conditions.

## Rules for filter and goto

Filter and goto definitions consist of question names and symbols only, no keywords (e.g. “goto”) are used.

- Symboles `( ) = < > @ | & : != <= >=`
- Filter `(AGE > 20) & (SEX = 1)` means: this question is asked if “age” is greater than 20 and “sex” is 1
- Goto `(2 @ TARGET)` means: if the answer to the current question is 2 then go to question “target”
- Refer to items using the colon as a seperator, e.g. `(PSOR:2 = 3)`.
- Value lists and ranges: `(x = 1:3)` is equal to `(x = 1,2,3)` is equal to `(x = 1) | (x = 2) | (x = 3)`

- Gotos only* evaluate the answer of this item and direct to the appropriate next answer.
- Room for improvement (e.g. loops), but works!

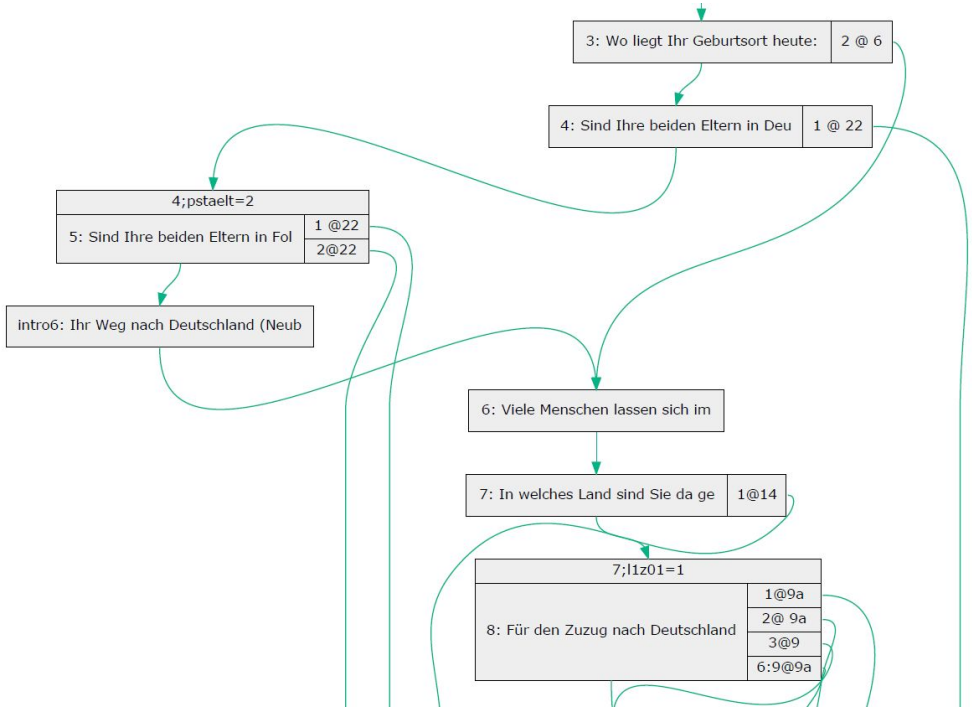
* have to update

<< Screenshot:

http://ddionrails.org/imports/questions_csv.html

# Use Case: Questionnaire Documentation

## Example: Visualize routing




- Flow chart, algorithmic derived from DDIonRails metadata
- Filters displayed
- Gotos parsed
- Layout/rendering by Graphviz

How is filter/goto-approach connected with visualisation?

# Make information re-usable and deal with changes

Re-use

- Means: Combine parts of a question and give them an identificator, which has to be used if the question appears again.
- Tracks permanence.
- Helps to limit amount of information, which has to be managed (entered, translated).
- Makes things more complicated: one more relation.
- Agency needed: assign IDs, ensure integrity, supervise corrections (internal question bank)

Link over time

- Same methods like those presented for variables
- Comparison seems to be more appropriate

Which parts of a question do you/ would you make re-useable?

How many resources do you have to track and describe changes?

Thank you for your attention.

**DIW Berlin — German Institute for Economic Research**
**(Deutsches Institut für Wirtschaftsforschung e.V. )**
Mohrenstraße 58, 10117 Berlin, Germany
www.diw.de

**Socio-Economic Panel (SOEP)**
www.soep.de
Knut Wenzig, kwenzig@diw.de

Mitglied der Leibniz-Gemeinschaft

# Reference and License

# Reference

This presentation is a major revision of:

Hebing, Marcel, & Wenzig, Knut (2016): Documenting Panel Data. Zenodo. doi:10.5281/zenodo.55613

# License

This presentation is offered under license CC BY-SA 4.0.

**The license does not apply to the following copyrighted material used in this prsentation:**

the logo of DIW Berlin/SOEP

the logo of the Leibniz Association