The DASISH Question Variable Data Base: Developments and Outlook

Håvard Venge Bakkmoen, Hilde Orten, Benjamin Beuster

NSD - Norwegian Social Science Data Services

Outline

- The Question Variable Data Base – purpose, aims and requirements
- QVDB – production pipeline, example use-cases
- Current structure and outlook for future developments
The Question Variable Data Base (QVDB)

• Purpose:
  - Searchable database with broad public profile with user access to survey questions in original languages, concepts, variables etc.

• Primary aims:
  - To serve business processes of the ESS and other surveys
  - Browsing possibilities for researchers and students
  - Interoperability with other systems and tools
  - Reusable model and code

• Possible users (same as for the QDDT):
  - European Social Survey (ESS) is the usecase
  - Other DASISH survey projects
  - Projects outside DASISH
  - Researchers and students
QVDB (and QDDT), a selection of the requirements

- DDI - Lifecycle based storage structure for metadata elements of high level of granularity
  - multilinguality
  - study-independent components
- DDI - Lifecycle and DDI-Codebook export and import possibilities to/from tools/web-services
- Communication between the three tools should be possible; minimal human interaction
- Support boolean field level search
- Core module based on DDI components; DDI profile; Resource Packages
- Open source system
- User access rights
QVDB in the archive production pipeline

- The QVDB is under development for use in the archive business processes at different steps in the pipeline:
  - Creating new variables
  - Creating data protocols and variable specifications for statistical packages
  - Documenting harmonisations etc.

- The QVDB will support import of questions from the QDDT and translations from the TMT

- This facilitates creating new variables as well as documenting them
Use case for specifications: ESS production pipeline

- Data protocol development for data and para-data.
- Variable definitions in statistical packages

Consultation processes country-specific background variables for harmonisation purposes.

Preparation of National Technical Summary

Data preparations e.g.
- Coding to international standards
- Harmonisation of country-specific variables
- Checks and editing of data into international standard variables.
- Anonymisation/data disclosure issues

Other fieldwork preparations, e.g. interviewer training.

Evaluations, specify needs for new ESS round.

CST, with input from other actors

Sampling methods/procedures

NTs, GESIS (Sampling team)

Source Questionnaire (and showcard) design and development

City and module teams
Comments from NCs and CST

Prepare archive processing
- Check and editing programmes
- working routines
- data processing guidelines

Archive

Data files deposit
(data and paradata files)

NTs

Deposit of National Technical Summary incl. appendices containing metadata related to background variables harmonisation, other technical

City, Archive, NTs

Data processing and quality checks/evaluations, editing and coding into standard variables (include data processing reports, various editions, expert comments)

Archive, NTs

Production of final country files

- Integration of country files
- Production of derived measures
- Production of multiple data products

NTs, GESIS

Metadata preparation

Archive

Data and metadata dissemination

CST, Archive NTs, users

Evaluation of data and metadata
- feedback

NTs
QVDB: possible use case – archive processes:
Create new variables for a questionnaire module for a new ESS round

1) View source questions in the module of the current round (QS R5)
2) View comparison between question list for module in the current round and that of the former round (Comparison QS R4 and QS R5)

a) No changes in question list:
   • View corresponding variable list and included elements
   • Reuse variable list used in former round (VS R4) or update if necessary

b) Questions deleted from list:
   • Delete and update corresponding variables from variable list used in former round (VS R4)

c) Questions added to list:
   • Check if the question has been used earlier and if so if reusable corresponding variables exist in the database.
   • Reuse, update or create new variable.

d) Questions changed:
   • Check earlier versions of the question and whether corresponding variables exist that could be reused.
   • Reuse, update or create new variable.

A new variable could have a based-on reference to a variable it is based on
QVDB: Aim - to support structuring of coding schemes and harmonisations
Example: ESS coding scheme for legal marital status

<table>
<thead>
<tr>
<th>Legal marital status categories for Iceland</th>
<th>Legal marital status ESS coding frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>Í hjónabandi</td>
<td>Legally married</td>
</tr>
<tr>
<td>Í staðfestri samvist</td>
<td>In a legally registered civil union</td>
</tr>
<tr>
<td>Skilin/n að borði og sæng</td>
<td>Legally separated</td>
</tr>
<tr>
<td>Skilin(n) að lögum/lögskilnaður/hjón askilnaður</td>
<td>Legally divorced / Civil union dissolved</td>
</tr>
<tr>
<td>Ekkja/ekkill</td>
<td>Widowed / Civil partner died</td>
</tr>
<tr>
<td>Ekkert af ofantöldu (hef ALDREI verið í hjónabandi)</td>
<td>None of these (NEVER married or in a legally registered civil union)</td>
</tr>
</tbody>
</table>

Categories have concepts

The word "Hjonaband" is like the word marriage in English. It describes the legal marital status of the two people. Two person in "Hjonaband". Í hjónabandi" refers to both same-sex and two-sex marriage.

Comparison: CategoryMap or grouping?

Corresponding categories
QVDB: possible use case – archive processes (2):
Development and reuse of categories from the ESS coding scheme for legal marital status in response domains and variable representations

1) Categories developed in the QVDB

Country-specific question, Iceland

2) Categories copied to TMT and QDDT and reused in questionnaire response domains

3) Categories reused in VariableRepresentations in QVDB

Legal marital status variable, Iceland

<table>
<thead>
<tr>
<th>MARSTIS</th>
<th>Categories for Iceland</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>í hjónabandi</td>
</tr>
<tr>
<td>2</td>
<td>í staðfestri samvið</td>
</tr>
<tr>
<td>3</td>
<td>Skilín/n að borði og sæng</td>
</tr>
<tr>
<td>4</td>
<td>Skilín(n) að líögum/lögskilnaður/hjónaskilnaður</td>
</tr>
<tr>
<td>5</td>
<td>Ekkja/ekkull</td>
</tr>
<tr>
<td>6</td>
<td>Ekkert af ofantöludu (hef ALDREI verið í hjónabandi)</td>
</tr>
<tr>
<td>77</td>
<td>Refusal</td>
</tr>
<tr>
<td>88</td>
<td>Don't know</td>
</tr>
<tr>
<td>99</td>
<td>No answer</td>
</tr>
</tbody>
</table>

Legal marital status variable, ESS

<table>
<thead>
<tr>
<th>MARSTIS</th>
<th>ESS coding frame</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Legally married</td>
</tr>
<tr>
<td>2</td>
<td>In a legally registered civil union</td>
</tr>
<tr>
<td>3</td>
<td>Legally separated</td>
</tr>
<tr>
<td>4</td>
<td>Legally divorced / Civil union dissolved</td>
</tr>
<tr>
<td>5</td>
<td>Widow/Civil partner died</td>
</tr>
<tr>
<td>6</td>
<td>None of these (NEVER married or in a legally registered civil union)</td>
</tr>
<tr>
<td>66</td>
<td>Not applicable</td>
</tr>
<tr>
<td>77</td>
<td>Refusal</td>
</tr>
<tr>
<td>88</td>
<td>Don’t know</td>
</tr>
<tr>
<td>99</td>
<td>No answer</td>
</tr>
</tbody>
</table>

Legal marital status categories for Iceland

- í hjónabandi
- í staðfestri samvið
- Skilín/n að borði og sæng
- Skilín(n) að líögum/lögskilnaður/hjónaskilnaður
- Ekkja/ekkull
- Ekkert af ofantöludu (hef ALDREI verið í hjónabandi)

Source question, ESS

F11 CARD 06: This question is about your legal marital status not about who you may or may not be living with. Which one of the descriptions on this card describes your legal marital status now?

CODE ONLY: PRIORITY CODE

- Legally married: 01
- In a legally registered civil union: 02
- Legally separated: 03
- Legally divorced / Civil union dissolved: 04
- Widow / Civil partner died: 05
- None of these (NEVER married or in a legally registered civil union): 06
- Not applicable: 66
- Refusal: 77
- Don’t know: 88
- No answer: 99
QVDB: Possible usecases - researchers and students

A researcher wishes to make an evaluation of the generated variables of a survey. He wishes to view all generated variables, as well as the syntaxes used to produce them. Find generated variables and the syntaxes they are based on. Make PDF of search results.

A student wishes to see which questions, variables and concepts have been used in a particular survey. Find questions, variables and concepts used in study, e.g. question lists, variable lists, concept lists, as well as information on single questions, variables and concepts and relations between them. Make PDF of search results.

A methodologist wishes to see how repeated metadata components in a survey, like questions and variables have developed over time. Find all versions of the metadata element of interest (question; variable) Views any comparisons made between different versions of the metadata element, e.g. variables. Make PDF of search results.

A researcher wishes to explore the coding standards used in a survey. Find and view the summary of DDI Resource Package. Find Resource Package used in survey. Create user defined enhanced report based on the content of the RP.

A researcher wishes to import questions in multiple languages into a different system than the QVDB for reuse. Find questions in multiple languages. Export DDI-XML instance or fragment from QVDB to different system.

A researcher wishes to import questions in multiple languages into a questionnaire delivery tool. Questions in multiple languages structured in detailed DDI is exported from the QVDB to questionnaire delivery tool.
QVDB Structure

• The QVDB builds on the database structure of the QDDT, but the QVDB also includes variables

• The conceptual model is based on DDI-Lifecycle
QVDB: Relationships between variables and questions

Survey

Variable

Source question

Translation eng_gb

Translation ger_de

Round 1

v1

Round 2

v2

Round 3

v3
QVDB – future outlook

Same as for the QDDT:
• Further development
• Real life use by ESS and SHARE
• Real life use by other survey projects
• Interoperability with other DDI-Lifecycle based and compatible survey tools

The QVDB:
• Include functionality to support further archive business steps, most importantly data curation

• The database could serve as a repository in a possible CESSDA DDI-Lifecycle based network (CESSDA portal)
Acknowledgements

Special thanks to Joachim Wackerow for consultancy

Thank you for your attention!

havard.bakkmoen@nsd.uib.no
hilde.orten@nsd.uib.no
benjamin.beuster@nsd.uib.no