DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Metadata across the survey lifecycle:
A common metadata understanding for the three DASISH survey tools

Hilde Orten, Knut Kalgraff Skjåk

NSD - Norwegian Social Science Data Services

Taina Jääskeläinen

FSD - Finnish Social Science Data Archive

Brita Dorer

GESIS - Leibniz Institute for the Social Sciences

Joachim Wackerow

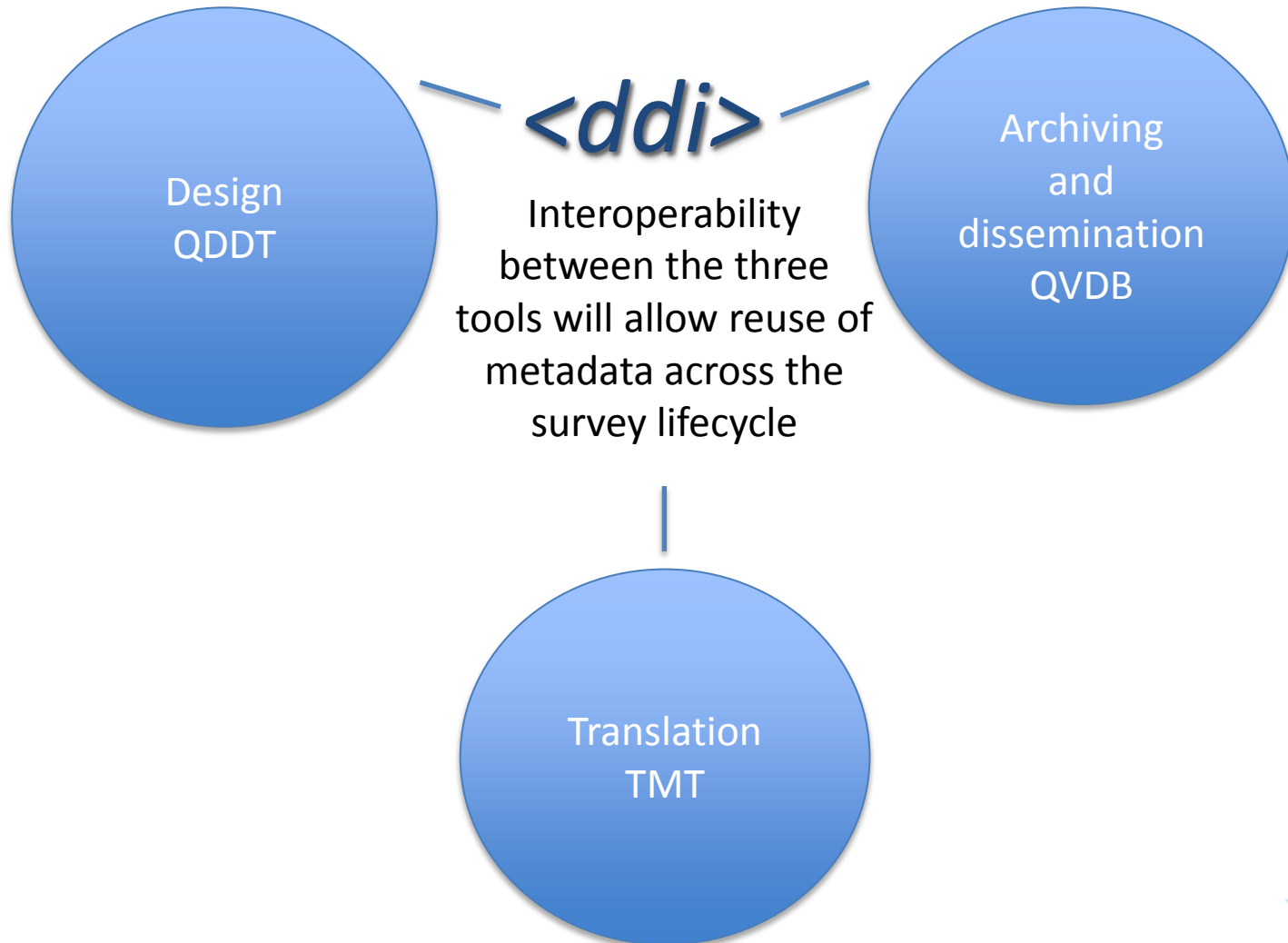*Consultant*

SEVENTH FRAMEWORK PROGRAMME

Outline

- The three DASISH task 3.2 survey tools and interoperability

- The metadata standard DDI

- A common metadata understanding for the three tools

DASISH software tools for the survey lifecycle

- Three software tools for surveys over under development:

  a) The Questionnaire Design and Development Tool (QDDT)
  b) the Question Variable Data Base (QVDB)
  c) the Translation Management Tool (TMT)

- The tools are under development as individual tools

- Goal: The tools should be able to communicate with each other so that metadata elements developed at a certain stage of the survey business process can be re-used at later steps

The DASISH tools and interoperability

Design QDDT

<ddi>

Interoperability between the three tools will allow reuse of metadata across the survey lifecycle

Archiving and dissemination QVDB

Translation TMT

What is DDI?

- The Data Documentation Initiative (DDI) is an international standard for describing data from the social, behavioral, and economic science

- DDI metadata specification supports the entire research data life cycle

- DDI metadata enables data conceptualization, collection, processing, distribution, discovery, analysis, repurposing, and archiving

- Facilitates export and import of metadata between software tools

DDI - Primary Benefits

- Rich content (currently over 800 items)

- Metadata reuse across the life cycle

- Machine-actionability

- Data management and curation

- Support for longitudinal data and comparison

- Support for preservation and platform-independent software

- Support for a global network

DDI – Example content

- Concepts ("terms")

- Studies ("surveys", "collections", "censuses", "experiments", etc.)

- Survey instruments ("questionnaire", "form")

- Questions ("observations")

- Responses

- Variables ("data elements", "columns")

- Codes & categories ("classifications", "codelists")

- Universes ("populations", "samples")

- Data files ("data sets", "databases")

DDI – Re-use Across the Lifecycle

Examples of metadata re-use across the survey lifecycle:

- Responses may use the same categories and codes which the variables use

- Multiple waves of a study may re-use concepts, questions, responses, variables, categories, codes, survey instruments, etc. from earlier waves

- Different studies may re-use questions from another survey etc.

# A common metadata understanding for the three DASISH tools, based on DDI

Requirements:

1. Which metadata elements will be used in the transfer between the three tools?

2. The direction for the flow of metadata elements between the three tools, as well as the steps in the work process at which metadata components are exchanged.

3. Mapping between the metadata elements and DDI.

4. A common identification and versioning system.

5. How the exchange of DDI metadata takes place

6. Administrative ownership of metadata

Metadata elements for transfer between the tools

- Metadata for transfer between the tools are typically questionnaire related elements that also are relevant for translation.

- Examples:

  - Questionnaire modules as in the instrument

  - Question items

  - Instructions

  - Categories and other response options

  - Complex question structures like grids

Possible metadata flow between the tools between milestones, generic example
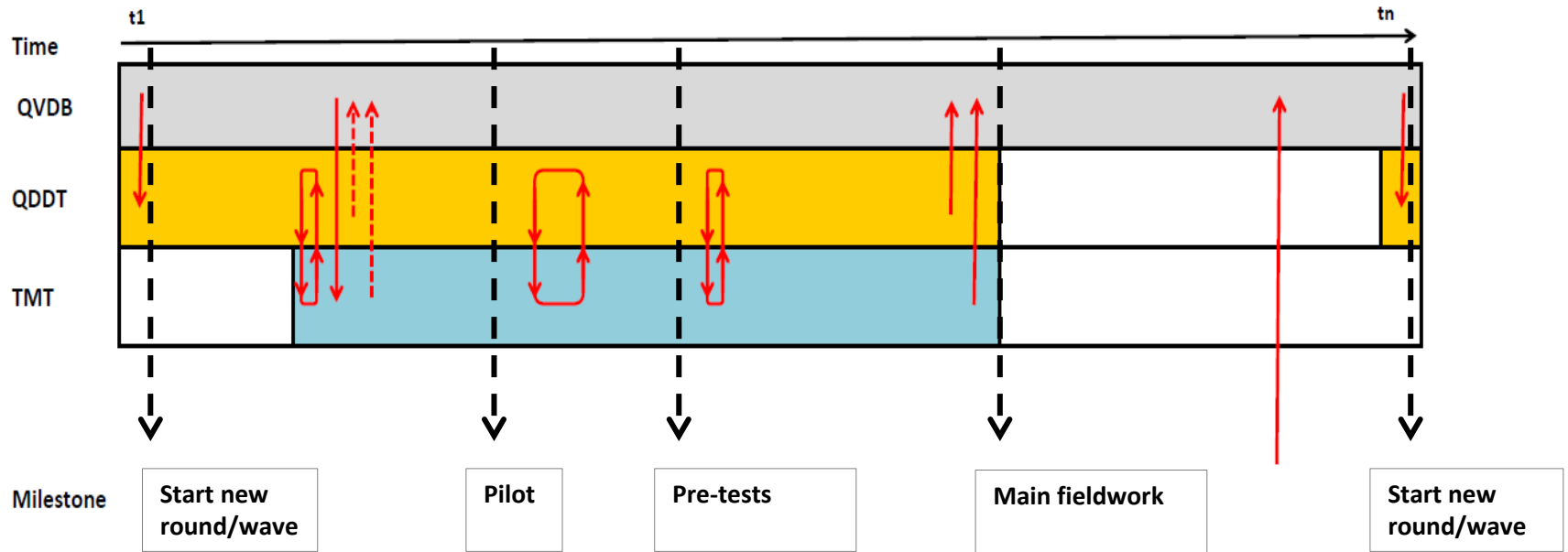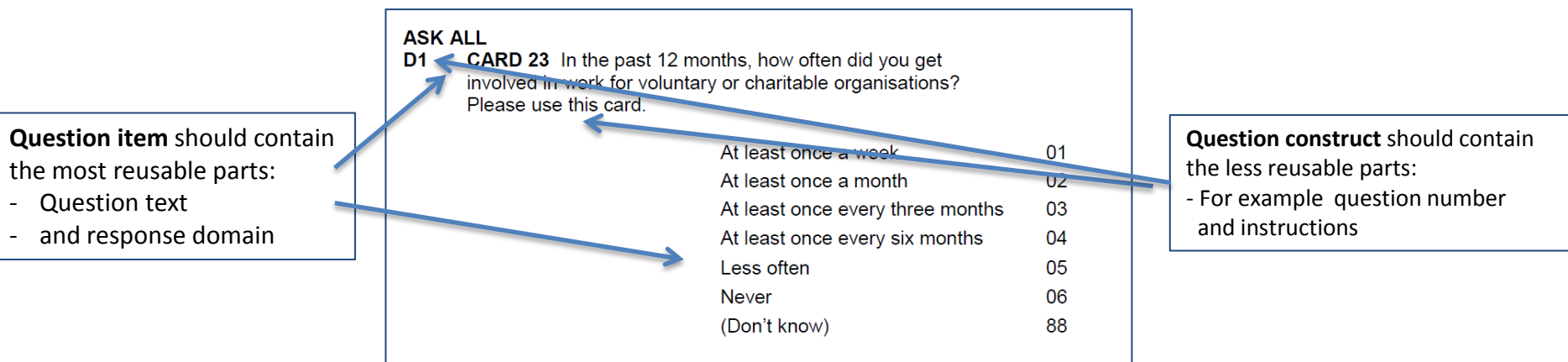
Structuring of metadata elements in DDI with reuse in mind

- Metadata elements like questions and concepts need to be mapped to DDI elements

- This should be done in a way that facilitates reuse of the elements within surveys over time as well as between surveys

Example: Question

- A question contains more or less reusable components

- More reusable and less reusable components should be structured in different

  DDI elements

Question item should contain the most reusable parts:
- Question text
- and response domain

ASK ALL
D1   CARD 23 In the past 12 months, how often did you get involved in work for voluntary or charitable organisations? Please use this card.

At least once a week	01
At least once a month	02
At least once every three months	03
At least once every six months	04
Less often	05
Never	06
(Don't know)	88

Question construct should contain the less reusable parts:
- For example question number and instructions

Versioning of metadata elements

- Main purposes of business versioning:

  - Enable humans and machines to distinguish changes in metadata elements that are important from those that are not so important

  - Track provenance

- Decisions regarding business versioning is purely driven by the decisions of the user

  - The user should assign a new version on the appropriate level according to the versioning policy, according to some versioning rationale decided by the versioning policy

Versioning policy

- DDI will be used for the business versioning

- All versionable DDI elements that are used should be versioned

- Two versioning levels will be used: Major and Minor

- Keeping versioning simple:

  • Minor change is changes in punctuation, spacing, capitalization or spelling, and other typographical and orthographical or minor wording changes that do not change the meaning.

  • All other changes are Major changes

Documenting the version rationale

- A rationale for the versioning should be described in a way that could be understood by machines as well as by humans

- Controlled vocabularies that describe the different types of changes to a metadata object will be used for this purpose

- Goal:
  - The systems should work so that a minor or major version is assigned automatically when a term from the controlled vocabulary is chosen by the user

- A description of the version rationale should also be possible

Documenting publication status

- It should also be possible for humans and machines to be able to distinguish between metadata elements that are not published, as well as between internal and external publications

- Controlled vocabularies for publication status are developed for this purpose

Example terms, publication status CV for the QDDT:

- Development, not published
- Internal publication, available for the questionnaire module design team
- Internal publication, available to 'export' to SQP for coding
- Internal publication, available to national teams to view
- Internal publication, available to 'export' to TMT for translation
- Internal publication, finalized items available to 'export' to the QVDB
- External publication, export to public

Versioning of translations

Example of parallel translation tracks by language (German) and country (Germany, Switzerland, Austria):
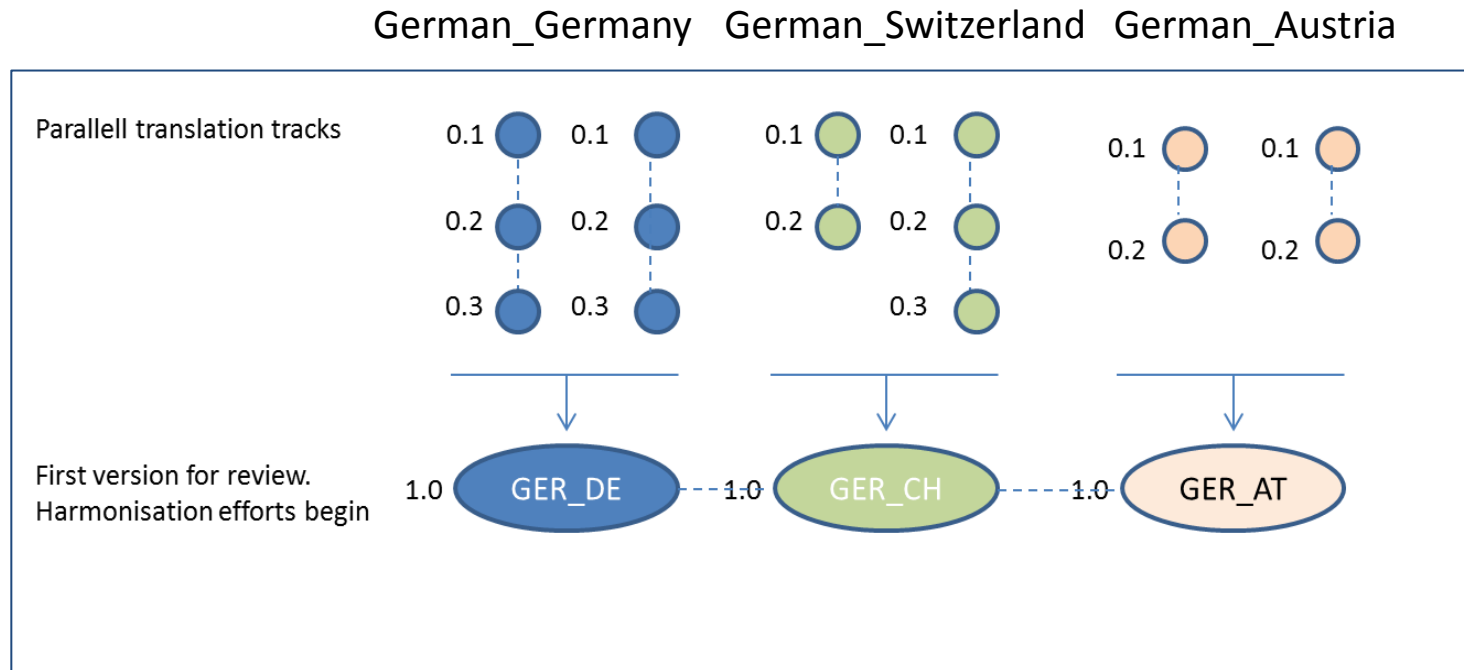
- Translations are versioned by county and language

- Reuse of metadata elements should be allowed within and between countries

- International administrator coordinates the versioning of shared languages

Administrative ownership of metadata

Proposed DDI domain names for the DASISH surveys:

- European Social Survey (ESS):

    int.esseric

- Survey of Health, Ageing and Retirement in Europe (SHARE):

    int.shareeric

Thank you very much for your attention!

hilde.orten@nsd.uib.no
knut.skjak@nsd.uib.no
taina.jaaskelainen@uta.fi
brita.dorer@gesis.org
joachim.wackerow@t-online.de

SEVENTH FRAMEWORK PROGRAMME