A Common Metadata Model for the Three DASISH Tools

The three tools of DASISH task 3.2

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Three software tools are under development under the Data Service Infrastructure for the Social Sciences and Humanities (DASISH) project, Work Package 3.2 are:

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

These tools are currently under development as individual tools that should be able to communicate with each other.

Interoperability between the tools using DDI

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

- Interoperability between the three tools will allow reuse of metadata across survey lifecycle business processes
- Will facilitate transfer of metadata developed in the QDDT and the TMT to the QVDB for storage of copies, search and reuse by a broad public
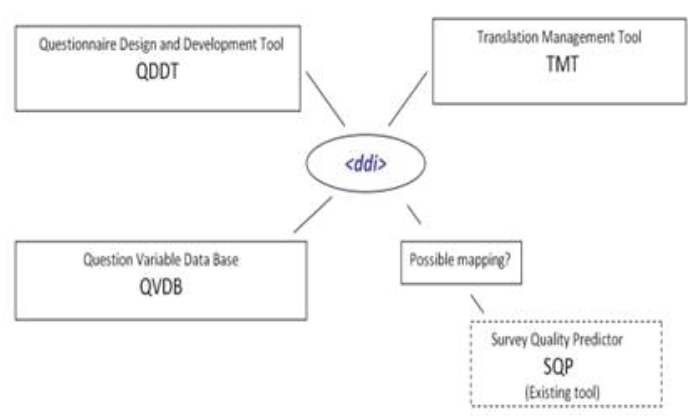- A common metadata understanding of the three tools facilitates interoperability

A common metadata understanding for the three tools

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Requirements for a common, DDI-L based metadata model*

- Which metadata elements will be used in the transfer between the three tools?
- Mapping between the metadata elements and the DDI
- The direction for the flow of metadata elements between the three tools, as well as the steps in the work processes at which metadata components are exchanged
- Administrative ownership of metadata
- A common identification and versioning system, including a versioning policy
- How the exchange of DDI metadata takes place, which type of DDI instances or fragments that will be transported, and which type of web service will be used

* The requirements were also presented in EDDI 2013

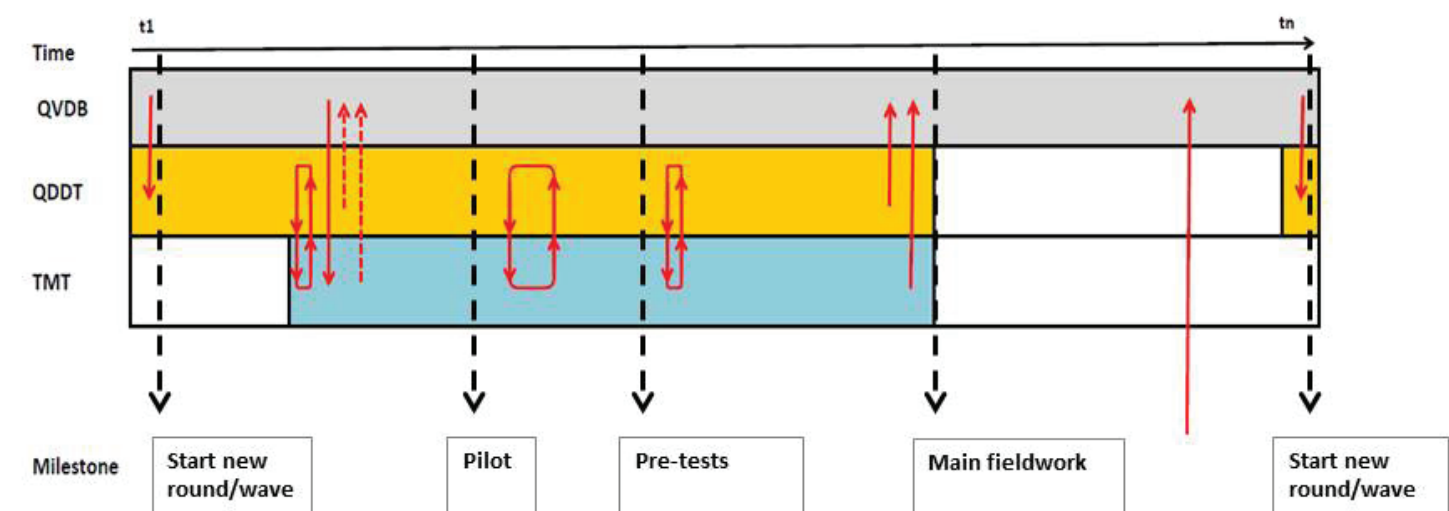## Which metadata elements are developed, maintained and stored in which tool (high level)?

	Development and maintenance (storing, and updating - add, change, delete)	Storing of copies
Concept and concept hierarchies	QDDT	QVDB, TMT
Questions and questionnaire related metadata elements for source questionnaires/modules	QDDT	QVDB, TMT
Translations of questionnaire related metadata elements	TMT	QVDB
Represented (reusable) variable Variables	QVDB	
Classification or coding schemes/ classification code-lists	QVDB	QDDT, TMT

Possible metadata flow between the tools between milestones, generic example

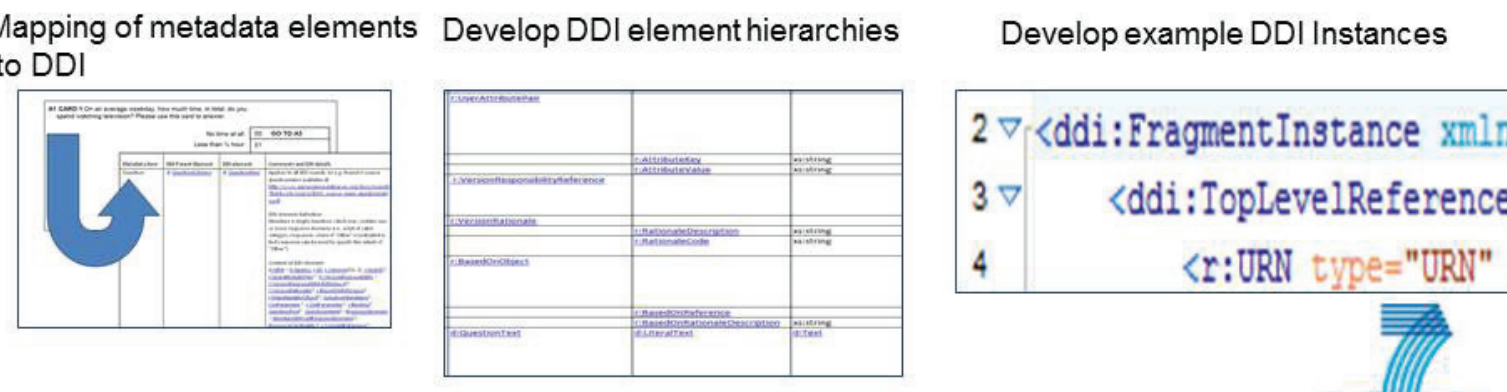DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Mapping to DDI – structuring of metadata elements

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

- Agreed exchange format: DDI 3.2
- Current focus: DDI profile for the metadata exchange between the tools
- Next step: DDI profiles for each of the three tools

Mapping to DDI – structuring of metadata elements (2)

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

How we work:

- Work out ideas and policies
- Meetings
- Discussions
- Expert consultancy
- DDI work, tasks below:

Mapping of metadata elements to DDI | Develop DDI element hierarchies | Develop example DDI Instances

```
2 <ddi:FragmentInstance xmln
3    <ddi:TopLevelReference
4       <r:URN type="URN"
```

Mapping to DDI – structuring of metadata elements (3)

Reusability:
Aim to develop and maintain reusable variants of questions and variables

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

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

Issue:
How to handle reusability in relation to complex question structures?

Mapping to DDI – structuring of metadata elements (4)

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

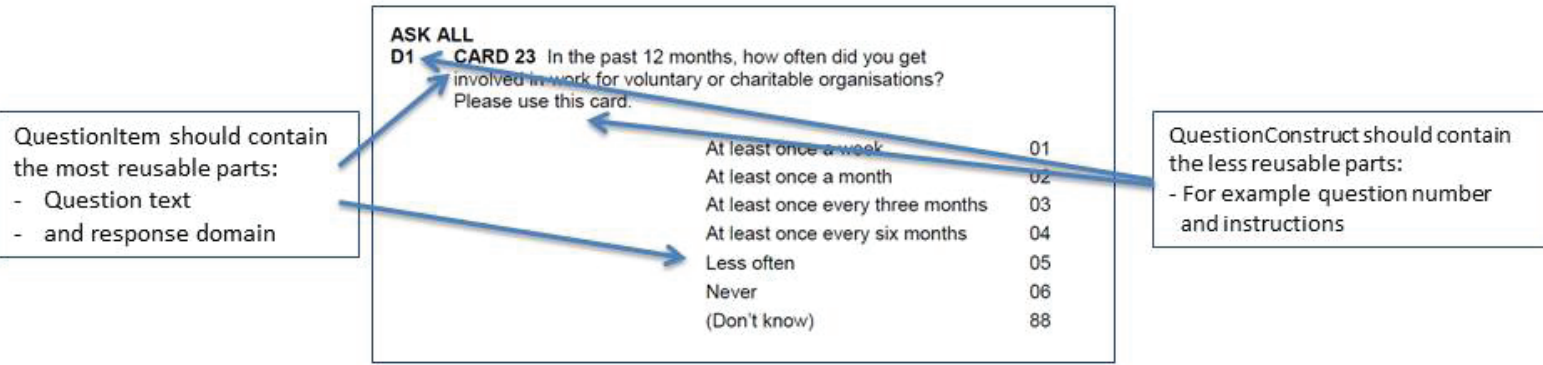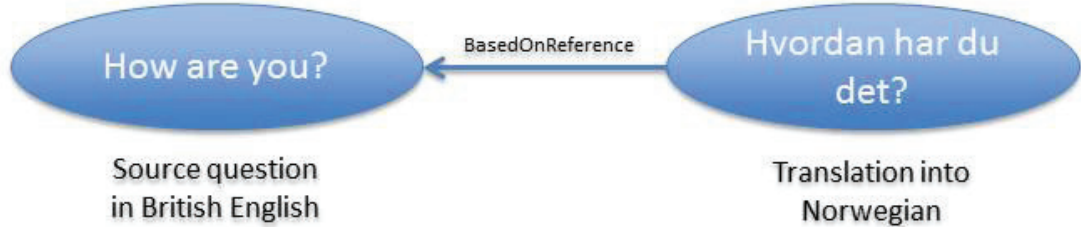Handling of translations of a source or master question:

- A translation can change across waves even if the source questionnaire remains the same
- Therefore source or master questions need to be treated as separate objects
- Translations should have a BasedOn reference to the source or master question it is translated from
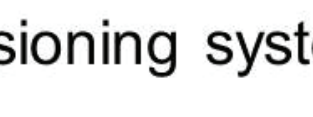

Source question in British English — Translation into Norwegian

Identification system

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

Identification system:

- DDI 3.2 Canonical URN will be used for identification:

  urn:ddi:agency:ID:Version

- The ID of the canonical URN will be a Universally Unique Identifier (UUID).
- Local identifiers can be stored as UserID in DDI

Versioning system – business versioning

- DDI is used for the business versioning
- The system should allow three versioning levels
- 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
- VersionRationale in DDI will be used to document version changes
  - RationaleCodes for describing different types of changes to a metadata object will be developed.
  - Goal: The user chooses a RationaleCode from a list to indicate what kind of change has been made, and the system automatically assigns either a major or minor change in the version number, according to the code chosen.
  - RationaleDescription will be allowed in the tool, but usage will be optional
- LateBound references should be allowed in the tool. Usage should be optional

Versioning system – business versioning (2)

- Each business version which should be made available to partners or to the general public should be flagged by the DDI attribute 'isPublished'
- UserAttributePair in DDI should additionally be used to distinguish between elements that are not published, as well as between internal and external publications

UserAttributePair usage:
Example controlled vocabulary for publication status for the QDDT

Code values for QDDT (ESS and possibly other surveys)

Suggested values for the drop-down list:

Development.NotPublished
InternalPublication.AvailableForDesignTeam
InternalPublication.ExportToQDDT
InternalPublication.AvailableForNationalTeams
InternalPublication.ExportToTMT
ExternalPublication.ExportToQVDB
ExternalPublication.ExportToPublic

Versioning policy:

DASISH — DATA SERVICE INFRASTRUCTURE FOR THE SOCIAL SCIENCES AND HUMANITIES

- All versionable DDI elements that are used should be versioned
- Business versioning should start when an object is available for search and archive
- Two versioning levels will be used: Major and Minor
  Minor change is typos or minor wording changes that do not change the meaning.
  All other changes are Major changes
- Batch versioning: User can choose two or more RationaleCodes, if more than one type of change is made to the element at one go, to describe all changes
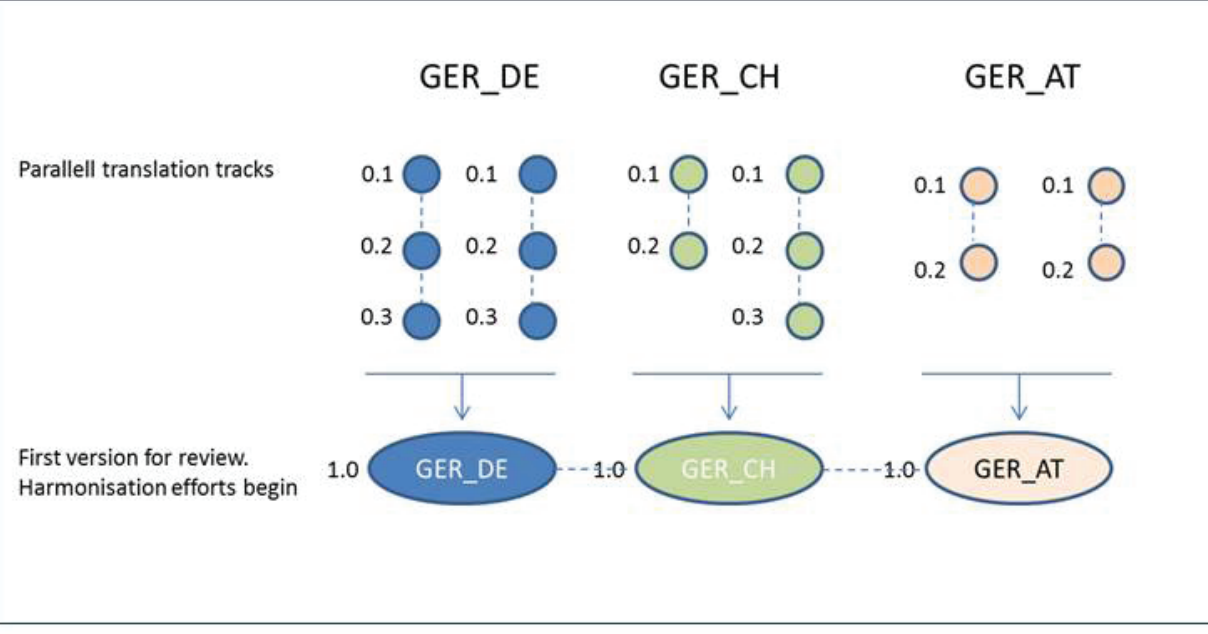
Versioning policy (2): Translations

Example of parallel translation tracks by language (German) and country (Germany, Switzerland, Austria)

- Translations are versioned by country and language
- Reuse of metadata elements should be allowed within and between countries
- International administrator coordinates the versioning of shared languages

Metadata exchange between the tools

- Full questionnaire modules as well as single questions will be exchanged between the tools
- Tools should be able to communicate via web services
  Goal: To develop REST API web services for the communication between the tools
- Possible further work: Communication between the three tools and SQP

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**