

Describing Qualitative
Surveys at beQuali: a
Crosswalk between
Documentation, Archiving
and Dissemination
Metadata

EDDI13 - 5th Annual European DDI User Conference
4 December 2013

Sarah Cadorel, Guillaume Garcia, Julien Rault
Paris

The project in a nutshell

- Who
 - Researchers
 - Students
- Why
 - Secondary analysis
 - Teaching (methods)
- What
 - Social Sciences : sociology, political sciences, anthropology mainly
 - Material originally digital or digitized
- How
 - No mandatory
 - Beyond raw data

The project in a nutshell

• Background

- 2005 : first thoughts
 - 2009 : feasibility study
 - 2011 : first web developments (DIME-SHS programme)
 - 2013 : dissemination website opening
early works about a self-deposit repository
- => We needed to improve the structure of our

The project in a nutshell

• Partnerships

- Réseau Quetelet & CESSDA
- Huma-Num (CNRS)
- Universities' archive services
- Digital long-term preservation (CINES)

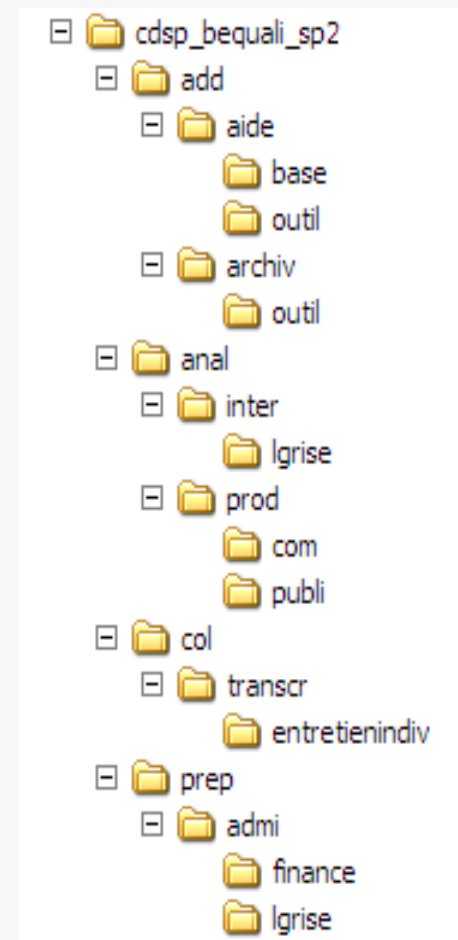
• Different metadata standards required

- We must use other standards besides DDI.
- We chose to use DDI 2.5 waiting to move on Qualitative data model for DDI at some point.

The project in a nutshell

• Our materials

- Our definition of a qualitative survey:
 - Preliminary documents (research programme, correspondence etc.)
 - Interviews/observations (transcriptions, notes etc.)
 - Analysis documents (analytical grid, pre-publication etc.)
- A complex corpus due to a large diversity of materials
 - Tree view ordering (by using practices and metadata from archive science).
 - Ergonomic dissemination (towards researchers by using in-house software)



Needs and challenges

- Our limits with DDI

- Qualitative survey description
 - Some specific fields are not handled in DDI 2.5
 - Ex : `locationofunitofobservation; studydate (#colldate #timeperiodcovered) ...`
- Materials (files) description
 - File level descriptive metadata
 - Ex: `continual numbering; content date; language...`
 - Technical (preservation) metadata
 - Ex: `filesize; compression; checksum...`
- Compound object aggregation
 - Tree view ordering
 - Versioning (content and format)

Needs and challenges

- Needs of interoperability
 - With archive services :
 - EAD import/export
 - SIP for long-term digital preservation
 - With digitization processes :
 - METS import
 - With data dissemination portals
 - DDI for Réseau Quételet
 - RDF for Huma-Num (on the way)
 - DC for others

Our strategy

- While waiting for qualitative data model for DDI to be fulfilled
 - move forward with the existing standards
 - try to map them with each others.
- In-house metadata format
 - combines the most relevant of each standard
 - Property names easy to understand by ITs
 - Based on JSON rather than XML for computational performance
 - remains internal & talks always with others within standards : we keep our mess at home.

Our metadata factory

- Crosswalk :

Level	DDI	EAD	METS	DC	PREMIS	SIP-CINES
Study	Widely used	Little used	Not used	Little used	Not used	Little used
Document	Little used	Widely used	Little used	Widely used	Widely used	Little used
Structure	Not used	Widely used	Widely used	Not used	Not used	Little used

■ Widely used ■ Little used □ Not used

Our metadata factory

- Example

```
"coverage_spatial_countries": ["FR"], # nation
"coverage_spatial_geographics": ["Paris"], # geogCover
"coverage_spatial_units": [ "7ème arrondissement", "18ème arrondissement"], # geogUnit
"coverage_temporal_begin": "1970", # timePrd@start
"coverage_temporal_end": "1980", # timePrd@end
"data_collection_context": ["Nonna Mayer a choisi d'interroger des petits commerçants
dans une première rue du 10ème arrondissement de Paris puis dans une deuxième rue du
7ème arrondissement de Paris. Elle a choisi de proposer systématiquement à tous les
commerçants de chaque rue de répondre à l'enquête. L'échantillon est composé des
ceux qui ont accepté de répondre aux entretiens."],
"data_collection_date_begin": "1972", # collDate
"data_collection_date_end": "1977", # collDate
"data_collection_methods": ["none directive interview"], # resInstru
"data_collection_modes": ["interview", "content analysis"], # collMode
"data_collection_samplings": ["Deux rues commerçantes à Paris, l'une dans un quartier
bourgeois (7ème) l'autre dans un quartier populaire (10ème). Tous les commerçants
ont été contactés, certains n'ont pas voulu répondre."], # sampProc
"data_collection_time_dimensions": ["one time interview"], # timeMeth
"data_languages": ["fr"] # ?
```

Upcoming

- Next steps :
 - to extend the implementation of this standard from the repository to the existing beQuali dissemination software,
 - to extend our metadata crosswalk to Qualitative data model for DDI as soon as possible.
- This is very much work in progress...
- Thanks for your comments and suggestions!

Questions?

www.bequali.fr

sarah.cadorel@sciencespo.fr

guillaume.garcia@sciencespo.fr

julien.rault@sciencespo.fr