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

EDDI13 — 5th Annual European DDI User Conference

4 December 2013

Paris

Sarah Cadorel, Guillaume Garcia, Julien Rault

beQuali

SciencesPo. cnrs

The project in a nutshell

- Who
  - Researchers
  - Students
- Why
  - Secondary analysis
  - Teaching (methods)
- What
  - Social Sciences : sociology, political sciences mainly
  - Originally digital or digitized material
- How
  - Not mandatory
  - Beyond raw data

The project in a nutshell

- Background
  - 2005: First conceptualized
  - 2009: Feasability study
  - 2011: First web application (DIME-SHS programme)
  - 2013: Dissemination website launched

    Early works about a self-deposit repository

  => **We needed to improve the structure of our metadata**

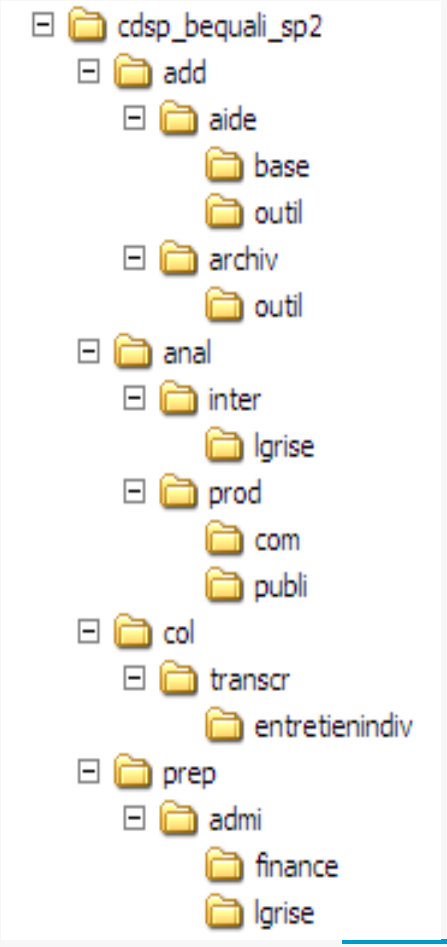
The project in a nutshell

- Partnerships
  - Data archives (Réseau Quetelet & CESSDA)
  - Digital Humanities (Huma-Num)
  - 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 planning to incorporate *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 Quetelet
    - RDF for Huma-Num (on the way)
    - DC for others

Our strategy

- While waiting for *Qualitative data model for DDI* to be completed
  - Move forward with the existing standards
  - Try to map them with each other.

- 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 and only communicates with others within the standards: we keep our mess at home.

Our metadata factory

- Crosswalk

Level	DDI	EAD	METS	DC	PREMIS	SIP-CINES
Study						
Document						
Structure						

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 10e arrondissement de Paris puis dans une deuxième rue du
    7e arrondissement de Paris. Elle a choisi de proposer aux commerçants de chaque rue
    de répondre à l'enquête. L'échantillon est composé de 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 (7e) l'autre dans un quartier populaire (10e). Tous les commerçants ont
    été contactés, certains n'ont pas voulu répondre."], # sampProc
"data_collection_time_dimensions": ["one time interview"], # timeMeth
"data_languages": ["fr"] # ?
```

Next steps

- Implement the standard from the repository within the existing beQuali dissemination software.
- Extend the metadata crosswalk to the 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