

Qualitative Data Exchange: Methods and Tools

Louise Corti

Abstract

This paper will present recent findings from a project that is exploring the feasibility of developing data exchange models and data conversion tools for primary research data collected in the course of qualitative research.

1. Summary

A standard format for representing richly encoded qualitative data is necessary because: it ensures consistency across datasets; it supports the development of common web-based publishing and search tools; and it facilitates annotated data interchange and comparison among data collections. Importantly, it should enable data and linked products to be imported and exported directly into and out of CAQDAS packages, avoiding the reliance on just a single product, and offering the opportunity to share analytic workings outside the confines of any particular software. The model and format have been requested by many other social science data archives round the world who are starting to collect qualitative research materials. Likewise there are researchers collecting multi media data in their own research who want to publish and showcases these data online, yet lack a consistent model of representing data that has a longer interpretive shelf life than basic html.

The project is developing, refining and testing models for data exchange for qualitative research data based on XML schema based on METS, while incorporating existing schema. These include: the Text Encoding Initiative (TEI), the Data Documentation Initiative (DDI) and Dublin Core and a system for identifying segments based on stand off annotation. The test data selected for this project are from the social sciences, but these formats are typically found across all domains of primary research. A key sample dataset has been taken the ESDS Qualidata archived project the Edwardians, that hold a range of data types, included data marked up in TEI.

The first step has been to liaise with and seek support from CAQDAS software vendors on a minimal data exchange model. A workable data exchange scenario rests on them providing export to and import from a common intermediate format. At the 2007 CAQDAS conference held at the University of London, a meeting of the majority for the vendors was held at which an initial model and schema was proposed. This was based on what might be termed core CAQDAS functionality, or a set of common denominators that could be replicated in the DExT data model. DExT has termed them **Data Constructs**.

The intermediate standard at the time of writing is likely to be called DEx-XML and QDEx-XML for the qualitative model, with any tools based on the model being known as DExT. The

full METS profile is being termed DEx-METS. The project has defined five major constructs (functions) that are common to all or most of the CAQDAS packages:

SEGMENTS: Identify Subsets of the study (e.g. Text or Line selections)

CODES: Assign Values to a Subset of a study, eg a segment

HiCODES: Create a Value Hierarchy (e.g. Codes arranged in a coherent hierarchical structure)

FileCLASS: Create a File Hierarchy/file classification (e.g. Files arranged in a coherent hierarchical structure)

MEMOS: Assign Notes or Comments (to a segment or a code)

Once the CAQDAS data have been transformed from vendor XML via XSLT stylesheets to the QDExT-METS, an enhancement interface enables additional metadata will be added. This includes the capacity to identify all the related files in the whole 'study' (given the possibility of complex multi-media collections) and looking at how the files relate to one and other. A viewer that transforms the DEx-METS XML will enable key mark-up of the collection to be demonstrated, though a fully functional viewer or interactive tool will not be produced under this project. It is hoped that if the model and schema are adopted other projects will consider developing user-oriented tools.

The project has engaged the services of an XML consultant from the Open Data Foundation (ODaF) who has helped finalise the first draft of the data model, the XML schema and a data processing interface (based on JAVA).

DExT is also working with intermediate data model and stand for the common statistical packages, SPSS, SAS and STATA. All the models and tools being created under this project are open source under the terms and conditions of the JISC award. The model will be published on Gforce and any tools on Source Forge. A beta release should be available by Autumn 2007.

References

DEXT www.data-archive.ac.uk/dext/

DEXT WIKI dawiki.essex.ac.uk/DEXTWiki/

Open Data Foundation (ODaF) <http://www.opendatafoundation.org/>

METS www.loc.gov/standards/mets/

UK Data Archive [ww.data-archive.ac.uk](http://www.data-archive.ac.uk)

The Edwardians www.esds.ac.uk/qualidata/online/data/edwardians/

About the Author

Louise Corti

DEXT

UK Data Archive

University of Essex

Colchester CO4 3SQ

44 1206 872145

corti@essex.ac.uk

