Data Exchange Tools (DExT)

DExT PROJECT

- funded under the JISC Repositories and Preservation Programme

- small budget for one year – proof of concept

- developing, refining and testing models and data conversion tools for data exchange for primary research data collected in the course of qualitative research

- test data selected are from the social sciences (multimedia, linked, annotated data etc.), but these formats are typically found across all domains of primary research

1 November 2006 – 31 March 2008

WHICH XML SCHEMA?

The selected output format chosen for DExT is the Metadata Encoding and Transmission Standard (METS) which serves both to describe the structure of, and to package, all the files relating to a study

METS is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the Extensible Markup Language (XML) schema language.

METS Metadata Encoding & Transmission Standard
Official Web Site

The DExT-METS XML format and editing Graphical User Interface (GUI) do not attempt to store or replicate the extensive functions offered by the various CAQDAS programs.

METS enables pointers to existing XML schemas in use to describe a study, project, file, extract or annotation:

- Metadata Object Description Schema (MODS)
- Dublin Core (DC)
- Text Encoding initiative (TEI)
- Data Documentation Initiative (DDI)
- Synchronized Multimedia Integration Language (SMIL)
- any other schemas that are relevant

AN OPEN EXCHANGE FORMAT FOR DATA

A standard uniform format for richly encoding research and data is necessary because it:

- enables long-term preservation and re-use of metadata, data and annotation (software- and platform-independent formats)

- ensures consistency of presentation and description of data

- facilitates the conversion of data to and from common statistical and qualitative data analysis (CAQDAS) packages using an open archival format specification

- supports the development of common web-based publishing and search tools

- enables more precise searching/browsing of archived data beyond the collection-level descriptive record

- facilitates data interchange, sharing among dispersed collections and repositories (comparative analysis and e-science)

VALUE ADDED TO DATA DURING QUALITATIVE DATA ANALYSIS

CAQDAS stands for Computer Assisted Qualitative Data AnalysiS, term, introduced by Fielding and Lee in 1991

- there are a wide range of software now available that supports a variety of analytic styles in qualitative work

- most have been under development for many years

- Examples: Atlas-ti, QDAMiner, Nvivo, MAxQDA

These functions are typically conducted within a proprietary environment:

- relationships between study objects: audio recording, transcript, observation

- context and enrichment of the data and study: memos, notes, annotations, outputs, global context

- analytic products: codes, classifications, relationships, linkages

UK Data Archive

www.data-archive.ac.uk/dext

UK Data Archive
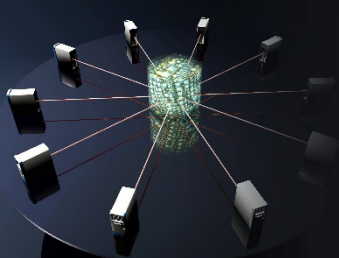
Data Exchange Tools (DExT)

METADATA STANDARDS - QUALITATIVE

The XML schema uses a reduced set of Text Encoding Initiative (TEI) elements:

- core tag set for transcription
- names, numbers, dates <persname>
- links and cross references <ref>
- notes and annotations <note>
- text structure <body>
- unique to spoken texts <kinesic>
- linking, segmentation and alignment <link>
- advanced pointing - XPointer framework
- text and AV synchronisation
- contextual information (participants, setting, text)

interview text with XML tags embedded

```
<u who="#interviewer" xml:id="u1">There's just
one or two factual things first of all do you mind
my asking how old you are?</u>

<u who="#subject" xml:id="u2">49.</u>

<u who="#interviewer" xml:id="u3">And what
schools did you go to?</u>

-<u who="#subject" xml:id="u4">

<orgName>King Street</orgName>
```

METADATA STANDARDS - STUDY

DDI 3.0 specification:

- ensures the availability of rich metadata that fully describe the data files

- captures the logical organisation of the information

- keeps track of the multiple physical instances in various formats

- manages multiple versions and is easily integrated into other DDI or XML compliant metadata management systems

COMMON CAQDAS CONSTRUCTS

Identify subsets of the study
e.g. text or line selections: SEGMENT

Assign values to a subset of a study
e.g. keywords or variables: CODE

Create a value hierarchy
e.g. keywords or codes arranged in a coherent hierarchical structure: HICODE

Create a file hierarchy
e.g. files arranged in a coherent hierarchical structure or classified: FILECLASS

Assign notes
e.g. comments or notes: MEMO

DExT PROGRESS

- comparison of relevant metadata/data schemas

- overview and use case analysis document

- Schemas – QuDEx and DEx-METS

- import of XML output from Atlas.ti and QDA Miner XML output into QuDEx

- UML model

- GUI functional specification for file import and metadata enrichment

- meeting with CAQDAS software vendors

- WIKI to discuss CAQDAS interchange standards

THE DExT PROJECT TEAM

Angad Bhat   Matthew Woollard   Louise Corti   Herve L'Hours

Open Data Foundation

supporting the vision of open and accessible data up and down the information chain

CONTACT

Louise Corti
UK Data Archive
University of Essex
Colchester, Essex CO4 3SQ

Email: corti@essex.ac.uk
Tel: +44 (0)1206 872145
URL: www.data-archive.ac.uk/dext