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 and survey 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

12 months duration
1 November 2006 – 31 October 2007

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

XML: enabling a standardised format for interview transcripts

METADATA STANDARDS - SURVEY

DDI 3.0 specification:

- ensures the availability of rich metadata that can fully describes the data files and dictionary

- 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

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

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

CAQDAS

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

- they use different methods to store links between annotated data and annotations

VALUE ADDED TO DATA DURING QUALITATIVE DATA ANALYSIS

These functions are typically conducted within a proprietary environment.

Relationships between study objects:
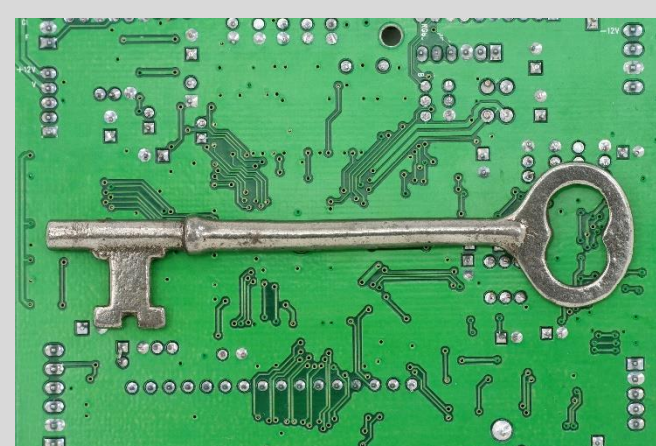
- audio recording
- transcript
- observation

Context and enrichment of the data and study

- memos
- notes
- annotations
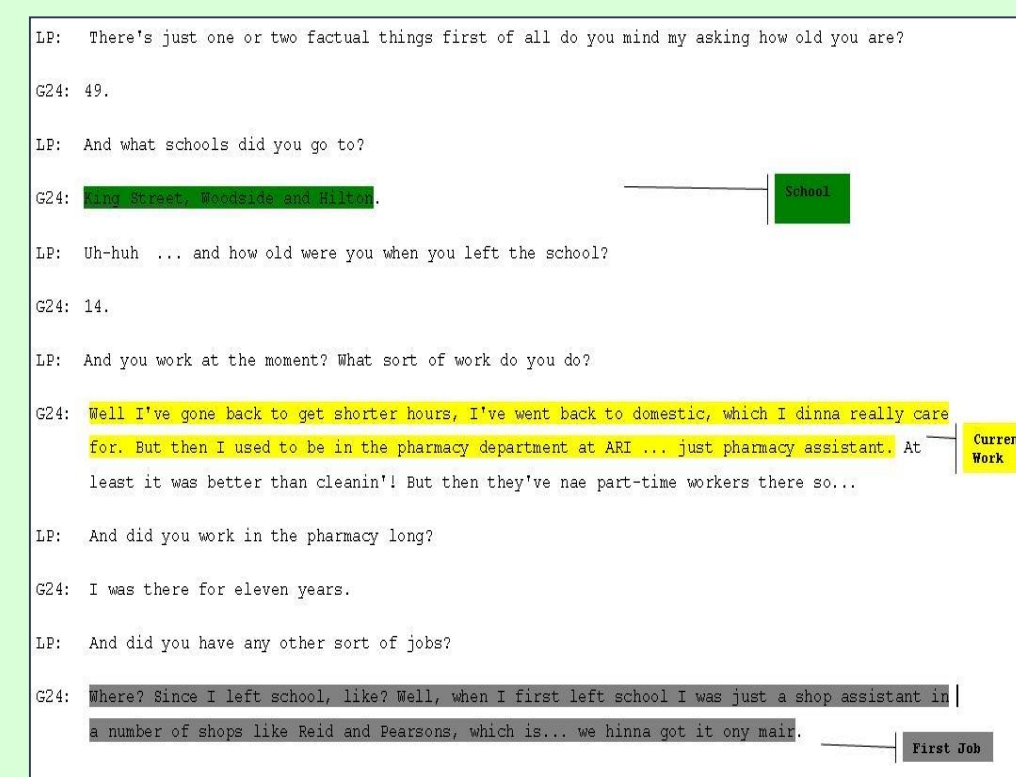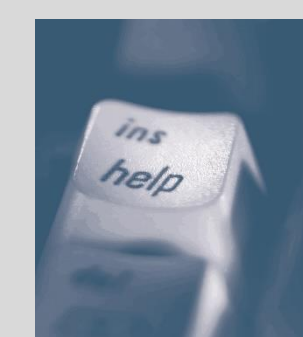- outputs
- global context

Analytic products

- codes
- classifications
- relationships
- linkages

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 Metadata Encoding & Transmission Standard *Official Web Site*

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

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

METS enables pointers to existing XML schema 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

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: HI CODE

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
- GUI functional specification for file conversion and metadata enrichment (DExT-METS)
- import from Atlas.ti and QDA Miner XML output into DExT-METS
- meeting with CAQDAS software vendors
- WIKI to discuss CAQDAS interchange standards
- GUI front end

SOLUTIONS

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

UK Data Archive

www.data-archive.ac.uk/dext

UK Data Archive