

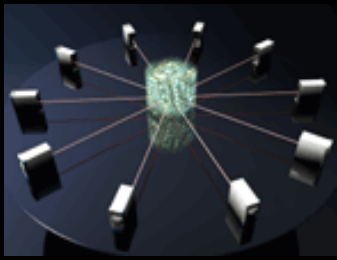
Data Exchange and Conversion Utilities and Tools (DExT)

Louise Corti, Angad Bhat, Herve L'Hours
UK Data Archive

CAQDAS Conference, April 2007

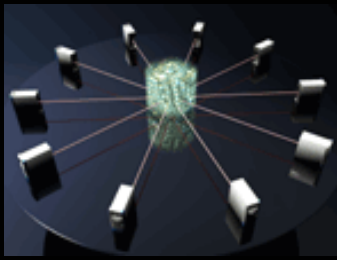
JISC





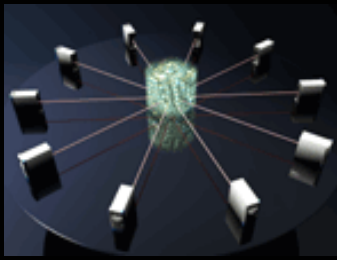
An exchange format for qualitative data

- Data exchange models and data conversion tools for primary research data collected in the course of qualitative research.
- A standard format for representing richly encoded qualitative data



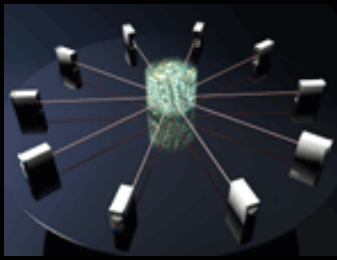
ESDS Qualidata

- national service led by the UK Data Archive (UKDA) –systematically archiving and enabling sharing of qualitative data since 1995
- focus is on acquiring digital data collections from purely qualitative and mixed methods contemporary research and from UK-based 'classic studies'
- facilitates the preservation of important large paper collections, and where appropriate, digitises samples of these collections.
- works closely with data creators (e.g academics) to ensure that high quality and well-documented qualitative data are produced
- offers user support and training to encourage professional researchers and research students alike to make full use of the rich sources of archived qualitative data



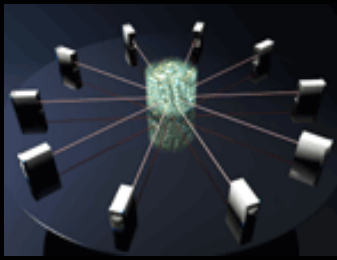
Access to data

- ESDS offers a resource discovery hub of some 4000 data collections
- some 160 qualitative research-based datasets
- developed an online data browsing service for texts (ESDS Qualidata Online)
- programme to extend and share common methods, standards and tools relating to this system
 - investigating new publishing forms: re-presentation of research outputs combined with data
 - investigating natural language processing, text mining and e-science applications to enable richer access to digital data banks



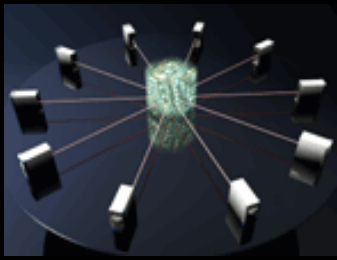
Applications of formats and standards for UKDA

- Long-term preservation requirements (software and platform independent formats)
- In-house toolsets for preparing qualitative data for multiple forms of dissemination
- Enable 'added-value' data to be retained – software-specific functionality
- Offers a standard for data creators to store and publish data in multiple formats eg common web-based publishing and search tools e.g ESDS Qualidata Online
- More precise searching/browsing of archived qualitative data beyond the catalogue record
- Facilitates annotated data exchange and data sharing across dispersed collections and repositories (comparative analysis and e-science)



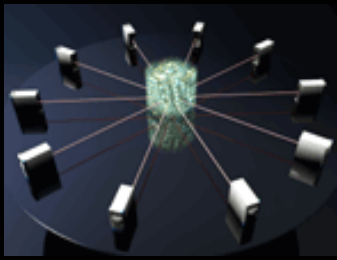
Added value

- Retain relationships between study objects
 - audio recording, transcript, observation
- Context – enrichment of the data and study
 - memos, notes, annotations, outputs, global context
- Analytic products – codes, classifications, relationships, linkages



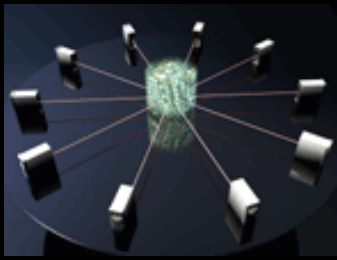
DExT Project

- JISC funded under Repositories Programme
- Small budget for one year – proof of concept
- Developing, refining and testing models for data exchange for qualitative research data based on XML/RDF schema
- 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



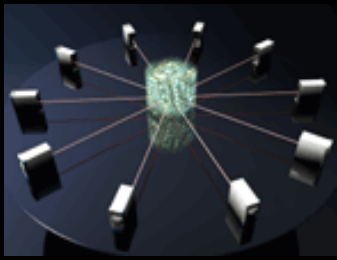
Which XML schema

- The selected output format chosen for DExT is the Metadata Encoding and Transmission Standard (METS) which serves to both describe the structure and to package all the files relating to a study
- METS – Metadata Encoding and Transmission Standard
 - is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language
 - The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation



METS

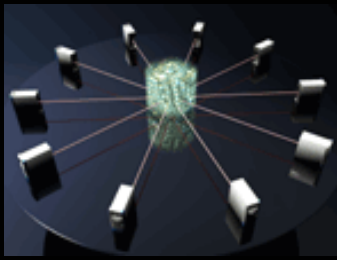
- Enables pointers to existing XML schema in use to describe a study, project, file, extract or say, annotation
 - Dublin Core
 - Text Encoding initiative (TEI)
 - Data Documentation Initiative (DDI)
 - QDIF
 - Triple S
 - Anything else relevant e.g ethno-methodological level annotation
- METS Navigator will allow browsing of all objects through a standard web browser



e.g TEI Schema

Qualidata 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)



Metadata for model transcript output

Study Name <titlStmt><titl>Mothers and daughters</titl></titlStmt>

Depositor <distStmt><depositr>Mildred
Blaxter</depositr></distStmt>

Interview number <intNum>4943int01</intNum>

Date of interview <intDate>3 May 1979</intDate>

Interview ID <persName>g24</persName>

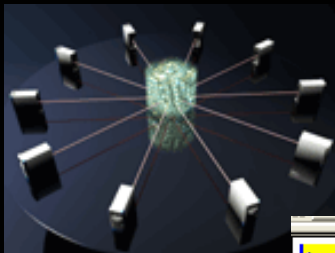
Date of birth <birth>1930</birth>

Gender <gender>Female</gender>

Occupation <occupation>pharmacy assistant</occupation>

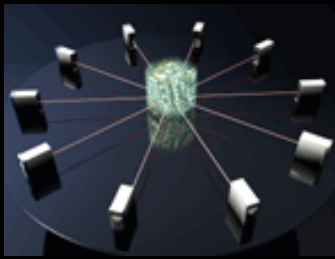
Geo region <geoRegion>Scotland</geoRegion>

Marital status <marStat>Married</marStat>



Transcript with XML mark-up

```
<xml>
<titlStmt><titl>Mothers and Daughters</titl></titlStmt>
<distStmt><depositr>Mildred Blaxter</depositr></distStmt>
<intNum>4943</intNum>
<intDate>3 May 1979</intDate>
<persName>q24</persName>
<birth>1930</birth>
<gender>Female</gender>
<occupation>Pharmacy assistant</occupation>
<geoRegion>Scotland</geoRegion>
<marStat>Married</marStat>
<!--comment -->
<u id="1" who="interviewer"> There's just one or two factual things first of all do you mind my asking how old you are?</u>
<u id="2" who="subject"> 49.</u>
<u id="3" who="interviewer"> And what schools did you go to?</u>
<u id="4" who="subject"> <orgName>King Street</orgName>, <orgName>Woodside</orgName> and <orgName>Hilton</orgName>.</u>
<u id="5" who="interviewer"> Uh-huh .. and how old were you when you left the school?</u>
<u id="6" who="subject"> 14.</u>
<u id="7" who="interviewer"> And you work at the moment? What sort of work do you do?</u>
<u id="8" who="subject"> Well I've gone back to get shorter hours, I've went back to domestic, which I dinna really care for. But then I used to be in the pharmacy department at <orgName>ARI</orgName> ... just <occupation>pharmacy assistant</occupation>. At least it was better than cleanin'! But then they've nae part-time workers there so..</u>
<u id="9" who="interviewer"> And did you work in the pharmacy long?</u>
<u id="10" who="subject"> I was there for eleven years.</u>
<u id="11" who="interviewer"> And did you have any other sort of jobs?</u>
<u id="12" who="subject"> Where? Since I left school, like? Well, when I first left school I was just a <occupation>shop assistant</occupation> in a number of shops like <orgName>Reid</orgName> and <orgName>Pearsons</orgName>, which is.. we hinna got it ony mair.</u>
<u id="13" who="interviewer"> Gone ..!</u>
<u id="14" who="subject"> <orgName>Marks and Spencers</orgName>, and the <orgName>Coop</orgName>. And that's about it.</u>
<u id="15" who="interviewer"> Uh-huh .. and then you got married and had your family. And when you went back to work, did you go into the pharmacy?</u>
```



XML: enabling a standardised format for interview transcripts

Study Name:: Mothers and Daughters
Depositor: Mildred Blaxter
Interviewer: Liz Patterson

Interview number: 4943int021.doc
Interviewer ID: g24
Date of interview: 3 May 1979

Information about interviewee

Date of birth: 1930
Gender: female
Marital status: married
Occupation: pharmacy assistant
Geographic region: Scotland

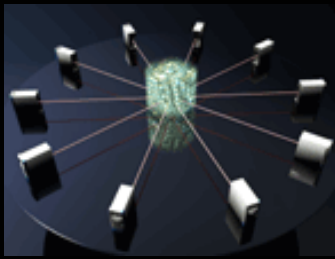
LP: There's just one or two factual things first of all do you mind my asking how old you are?

G24: 49.

LP: And what schools did you go to?

G24: King Street, Woodside and Hilton.

LP: Uh-huh .. and how old were you when you left the school?



XML and XSL: enabling web-enabled display, search and browse

ESDS | Home | A-Z index | Site map | Contact | Login | Search:

ESDS Qualidata

Economic and Social Data Service

[About](#) | [Data](#) | [Create/deposit](#) | [Online](#) | [Support](#) | [News](#) | [Events](#) | [Which service?](#)

[Introduction](#)
[Edwardians](#)
[Mothers and Daughters](#)
[100 Families](#)

Transcript

To view the interview summary, click the title. (Only available for Edwardians interviews at the moment)

g24

Born: 1930 Female
Occupational Class: Sales and customer service **Geographic Region:** Scotland

There's just one or two factual things first of all do you mind my asking how old you were when you left the school?

49.

And what schools did you go to?

King Street, Woodside and Hilton.

Uh-huh .. and how old were you when you left the school?

14.

And you work at the moment? What sort of work do you do?

Well I've gone back to get shorter hours, I've went back to domestic, which I dinna really

ID: g24
Born: 1930 Female
Occupational Class: Sales and customer service **Geographic Region:** Scotland
... and Hilton. Uh-huh. and how old were you when you left the **school?** 14. And you work at the moment? What sort of work do you ...

ID: g25
Born: 1932 Female
Occupational Class: Elementary **Geographic Region:** Scotland
... No, no, no. I was a shop assistant when I left **school**. And then we had our own shop down in Ashgrove there. my sister an ...

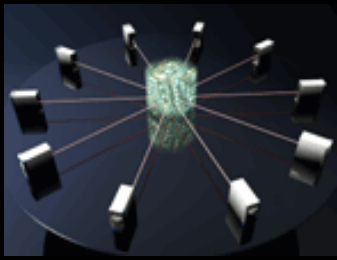
ID: g31
Born: 1929 Female
Occupational Class: Unemployed **Geographic Region:** Scotland
... , what sort of work did you do? When I first left the **school?** An upholstery sewer. I worked in Roberts in Union Street for 16 years, ...

ID: g32
Born: 1921 Female
Occupational Class: Unemployed **Geographic Region:** Scotland
... Gordon's College. And a shop assistant before that. And when you left **school?** Oh, when I left **school** I went to Broadford. When I left **school** ...

ID: g6
Born: 1921 Female
Occupational Class: Elementary **Geographic Region:** Scotland
... . I'm 57. And what schools did you go to? Old Aberdeen **School** for a start, and then Sunnybank. And how old were you when you left ...

ID: g7
Born: 1916 Female
Occupational Class: Unemployed **Geographic Region:** Scotland
... what schools did you go to yourself? Well, Porthill and the Middle **School**. Aye, Porthill's the Primary and the Middle's the Secondary. Yes. And how ...

You are here : ESDS Qualidata Online > Explore > IndexZoom > Interview transcripts: search results

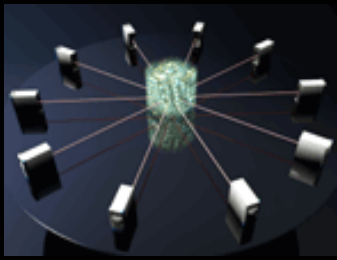


DExT progress so far



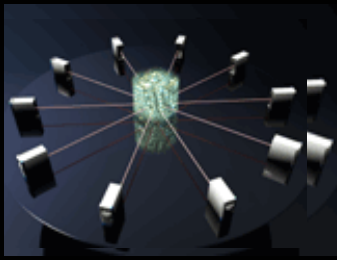
Produced:

- Comparison of relevant metadata/data schema
- Overview and Use Case Analysis document
- GUI Functional Specification for File Conversion & Metadata Enrichment (DExT-METS)
- Import from Atlas.ti and QDA Miner XML output into DExT-METS
- GUI front end
- Meeting with software vendors tonight for feedback



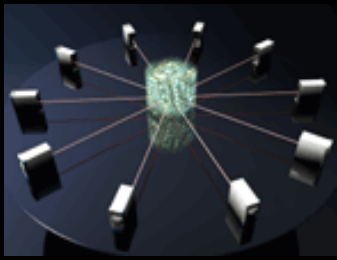
DExT-METS

- The DExT-METS XML format and editing GUI (DExT-METS Generator) do not attempt to store or replicate the extensive functions offered by the various CAQDAS programs
- The aim of DExT is to identify the common data constructs used across these proprietary formats and store them in a platform independent environment suitable for data interchange and long term preservation



Basic data constructs replicated in DExT

- Identify Subsets of the study
(e.g. Text or Line selections : [Quotation concepts](#))
- Assign Values to a Subset of a study
(e.g. Keywords or Variables : [Codes concept](#))
- Create a Value Hierarchy
(e.g. Keywords or Codes arranged in a coherent hierarchical structure : [SuperCodes concept](#))
- Create a File Hierarchy
(e.g. Files arranged in a coherent hierarchical structure : [Family concept](#))
- Assign Notes
(e.g. Comments or Notes: *Memos* concepts)



Identifying Subsets from the study (Quotation Concept)

LP: There's just one or two factual things first of all do you mind my asking how old you are?

G24: 49.

LP: And what schools did you go to?

G24: King Street, Woodside and Hilton.

LP: Uh-huh ... and how old were you when you left the school?

G24: 14.

LP: And you work at the moment? What sort of work do you do?

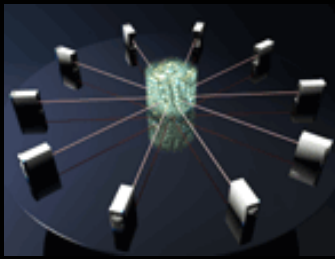
G24: Well I've gone back to get shorter hours, I've went back to domestic, which I dinna really care for. But then I used to be in the pharmacy department at ARI ... just pharmacy assistant. At least it was better than cleanin'! But then they've nae part-time workers there so...

LP: And did you work in the pharmacy long?

G24: I was there for eleven years.

LP: And did you have any other sort of jobs?

G24: Where? Since I left school, like? Well, when I first left school I was just a shop assistant in a number of shops like Reid and Pearsons, which is... we hinna got it only mair.



Assign Values to Subsets (Codes Concept)

LP: There's just one or two factual things first of all do you mind my asking how old you are?

G24: 49.

LP: And what schools did you go to?

G24: King Street, Woodside and Hilton.

School

LP: Uh-huh ... and how old were you when you left the school?

G24: 14.

LP: And you work at the moment? What sort of work do you do?

G24: Well I've gone back to get shorter hours, I've went back to domestic, which I dinna really care for. But then I used to be in the pharmacy department at ARI ... just pharmacy assistant. At least it was better than cleanin'! But then they've nae part-time workers there so...

Current
Work

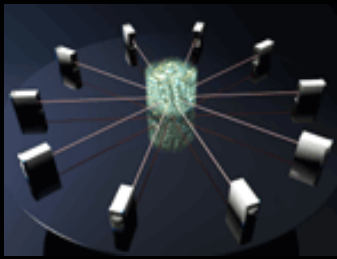
LP: And did you work in the pharmacy long?

G24: I was there for eleven years.

LP: And did you have any other sort of jobs?

G24: Where? Since I left school, like? Well, when I first left school I was just a shop assistant in a number of shops like Reid and Pearsons, which is... we hinna got it ony mair.

First Job



Create a value hierarchy (SuperCodes Concept)

LP: There's just one or two factual things first of all do you mind my asking how old you are?

G24: 49.

LP: And what schools did you go to?

G24: King Street, Woodside and Hiltor.

School

LP: Uh-huh ... and how old were you when you left the school?

G24: 14.

LP: And you work at the moment? What sort of work do you do?

G24: Well I've gone back to get shorter hours, I've went back to domestic, which I dinna really care for. But then I used to be in the pharmacy department at ARI ... just pharmacy assistant. At least it was better than cleanin'! But then they've nae part-time workers there so...

Current Work

LP: And did you work in the pharmacy long?

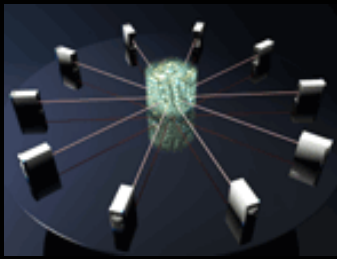
G24: I was there for eleven years.

LP: And did you have any other sort of jobs?

G24: Where? Since I left school, like? Well, when I first left school I was just a shop assistant in a number of shops like Reid and Pearsons, which is... we hinna got it ony main.

First Job

JOB



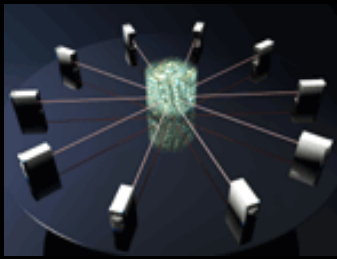
Create a file hierarchy (Family Concept)

 Male Family

-  Interview 1
-  Interview 2
-  Interview 3
-  Interview 4

 Female Family

-  Interview 1
-  Interview 3
-  Interview 2
-  Interview 4



DEXT-METS Generator GUI

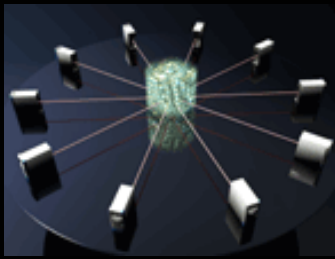
Please select a process from the list below

Main Menu

- ☒ Atlas.ti
- ☐ QDA Miner
- ☐ Manual DEXT-METS process

Proceed

[Next](#)



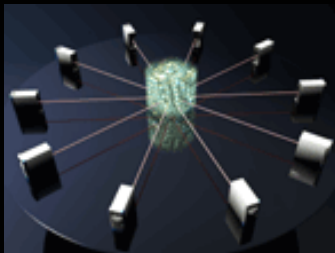
Atlas.ti conversion to DEX-T-METS

Atlas.ti conversion to DEX-T-METS

File Path

DEX-T-METS Editor: Once approved please click save

```
<?xml version='1.0' encoding='UTF-8'?>
<mets:mets
xsi:schemaLocation="http://www.loc.gov/METS/
http://www.loc.gov/standards/mets/mets.xsd"
xmlns:mets="http://www.loc.gov/METS/"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1"
xmlns:xlink="http://www.w3.org/1999/xlink"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:qdif="http://delacruz.csse.unimelb.edu.au/qdif"
xmlns:tei="http://www.tei-c.org/ns/1.0"
xmlns:teix="http://www.tei-c.org/ns/Examples"
xmlns:rng="http://relaxng.org/ns/structure/1.0">
<mets:metsHeader>2007-02-01T12:00:18
</mets:metsHeader></mets:mets>
```



Text Encoding Initiative for METS

TEI Header for METS file

Major Modules

- ☒ File Description
- ☒ Encoding Description
- ☒ Profile Description
- ☒ Revision History

Module Elements

File Description	Encoding Description	Profile Description	Revision History
<input checked="" type="checkbox"/> Title Statement <input checked="" type="checkbox"/> Publication Statement <input checked="" type="checkbox"/> Source Description <input checked="" type="checkbox"/> Notes Statement <input checked="" type="checkbox"/> Edition Statement <input checked="" type="checkbox"/> Series Statement <input checked="" type="checkbox"/> Extent	<input checked="" type="checkbox"/> Project Description <input checked="" type="checkbox"/> Sampling Description	<input checked="" type="checkbox"/> Participant Description <input checked="" type="checkbox"/> Setting Description	<input checked="" type="checkbox"/> Revision Description

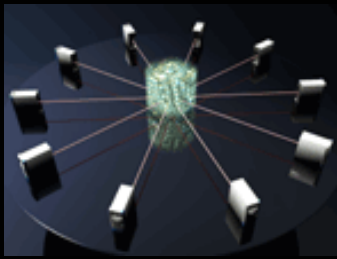
Sub-Elements and Attributes

Title Statement <input checked="" type="checkbox"/> Title <input checked="" type="checkbox"/> Sponsor <input checked="" type="checkbox"/> Funder <input checked="" type="checkbox"/> Responsibility Statement * <input type="checkbox"/> Title Level <input checked="" type="checkbox"/> Type	Responsibility Statement <input checked="" type="checkbox"/> Responsibility <input checked="" type="checkbox"/> Name <input checked="" type="checkbox"/> Person Name	Publication Statement <input checked="" type="checkbox"/> Publisher <input checked="" type="checkbox"/> Authority <input checked="" type="checkbox"/> Distributor <input checked="" type="checkbox"/> ID Number <input checked="" type="checkbox"/> Publication Place <input checked="" type="checkbox"/> Date <input checked="" type="checkbox"/> ID Number Type	Source Description <input checked="" type="checkbox"/> Bibliographic citation	Notes Statement <input checked="" type="checkbox"/> Notes	Extent <input checked="" type="checkbox"/> Extent	Series Statement <input checked="" type="checkbox"/> Title <input checked="" type="checkbox"/> ID Number <input checked="" type="checkbox"/> Responsibility Statement * <input type="checkbox"/> Title Level <input checked="" type="checkbox"/> Type <input checked="" type="checkbox"/> ID Number Type	Responsibility Statement <input checked="" type="checkbox"/> Responsibility <input checked="" type="checkbox"/> Name <input checked="" type="checkbox"/> Person Name
Participant Description <input checked="" type="checkbox"/> List Person <input checked="" type="checkbox"/> Person Group <input checked="" type="checkbox"/> Person * <input checked="" type="checkbox"/> Participant Links * <input type="checkbox"/> List Person Type <input type="checkbox"/> Person Group ID <input type="checkbox"/> Person ID <input type="checkbox"/> Role <input type="checkbox"/> Sex <input type="checkbox"/> Age	Person <input checked="" type="checkbox"/> Person Name <input checked="" type="checkbox"/> Birth <input checked="" type="checkbox"/> Place Name * <input checked="" type="checkbox"/> Death <input checked="" type="checkbox"/> Paragraph <input checked="" type="checkbox"/> Sex <input checked="" type="checkbox"/> Occupation <input checked="" type="checkbox"/> Person State * <input type="checkbox"/> Person Name ID <input type="checkbox"/> Birth Date <input type="checkbox"/> Death not Before <input type="checkbox"/> Death not After <input type="checkbox"/> Sex Value <input type="checkbox"/> Person state Type	Place Name <input checked="" type="checkbox"/> Settlement <input checked="" type="checkbox"/> Country <input checked="" type="checkbox"/> Person State <input checked="" type="checkbox"/> Paragraph <input type="checkbox"/> Settlement Type <input type="checkbox"/> Country Reg <input type="checkbox"/> Relation Type <input type="checkbox"/> Relation Name <input type="checkbox"/> Active <input type="checkbox"/> Passive	Setting Description <input checked="" type="checkbox"/> Paragraph <input checked="" type="checkbox"/> Setting * <input type="checkbox"/> Setting who	Setting <input checked="" type="checkbox"/> Name <input checked="" type="checkbox"/> Date <input checked="" type="checkbox"/> Locale * <input checked="" type="checkbox"/> Activity <input type="checkbox"/> Name Type <input type="checkbox"/> Date Value <input type="checkbox"/> Activity Value	Locale <input checked="" type="checkbox"/> Place Name	Project Description <input checked="" type="checkbox"/> Paragraph	Sampling Description <input checked="" type="checkbox"/> Paragraph
		Revision Description <input checked="" type="checkbox"/> Change * <input type="checkbox"/> Change Number	Change <input checked="" type="checkbox"/> Date <input checked="" type="checkbox"/> Responsibility Statement * <input checked="" type="checkbox"/> Item <input type="checkbox"/> Date Value <input type="checkbox"/> Date Certainty	Responsibility Statement <input checked="" type="checkbox"/> Responsibility <input checked="" type="checkbox"/> Name <input checked="" type="checkbox"/> Person Name			

<<Back

Save & Proceed

Next



METS File Section

File Section for the METS file

File Section

Text

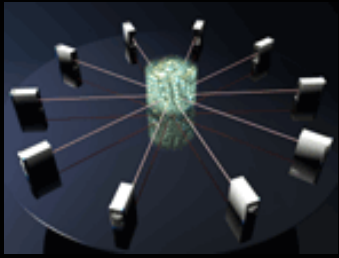
Xml

Audio

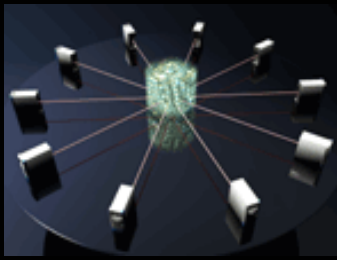
Video

Image

[Next](#)

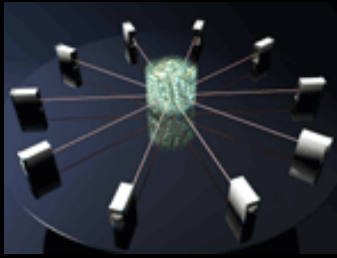


Some use cases



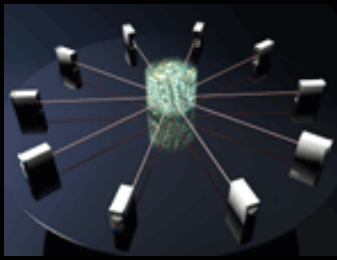
Preservation requirements

- Terms of the grant - all project output should be made available with preservation-level metadata. The most appropriate tool to manage the process would be the vendor's product which also has the capability to export to DExT-METS format
- The Researcher has met a requirement from the funding body with no additional expense of time or energy while ensuring the long term availability of both the vendor-specific and the platform independent versions of the study
- Depositor gains by having a nearly push-button solution to creating deposit-ready data, and UKDA saves on processing time



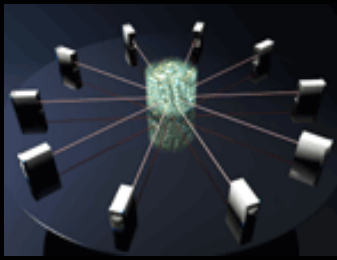
Vendor-Specific Functionality

- An extensive project developed in an environment completely reliant on Vendor Ones program would benefit from additional analysis using different functionality only available in Vendors Twos program
- Least-common-denominator model



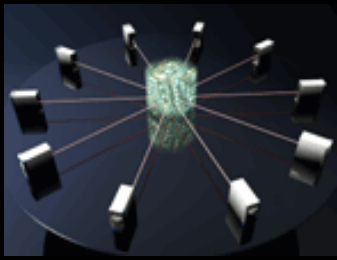
Analysis of Legacy Data

- Vast quantities of legacy data available from a past project would benefit from analysis using modern tools
- The original project relied on a proprietary tool which, while still in existence, is not backwards compatible with the relevant output. However, copies of the content were output in DExT-METS
- The core data of the historical project is still available and may be transformed into the latest version of the DExT-METS format and imported into modern compliant CAQDAS programs



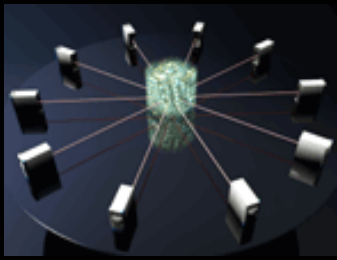
Vendor-Specific Markup via 3rd Party Tools

- An extensive collection of documents have received funding to make them available online to the wider academic community. In addition to conversion of the original content to html format all qualitative analysis has been output to DExT-METS format
- The developers of the web interface now have access to a fully documented open source format describing the structure and content of the study, facilitating the creation of a resource discovery framework.
- They also have access to a considerable body of work originally created with the vendors program to mark up the text which can be repurposed for display online



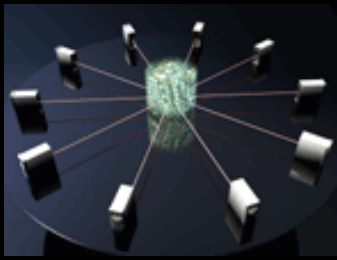
Metadata Enrichment of Resources

- An extensive qualitative study is not deemed suitable for ingest into repositories because of the proprietary nature of the analysis output and the absence of standard compliant descriptive and technical metadata accompanying the resource
- A Researcher exports the collection to DExT-METS for interoperability and uses the DExT-METS Generator to generate a standard TEI header and unqualified Dublin Core suitable for harvesting under OAI-PMH



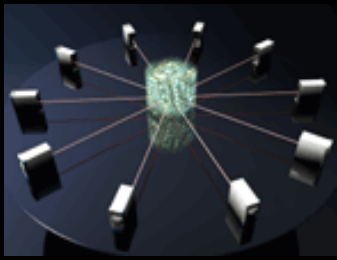
From Vendor-Specific to Vendor-Neutral

- The DExT project proof of concept work includes plans to convert Atlas.ti and QDA Miner (both available as XML exports) to a draft version of the DExT-METS format. In the future there are two possible mechanisms for the creation of vendor-neutral resources
 - 3rd party creation of tools to transform vendor XML output to DExT-METS
 - Vendor outputs directly to DExT-METS format



Assumptions for take-up

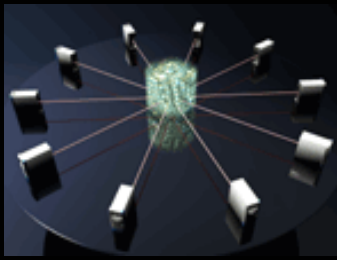
- Core data concepts can be exported to DExT-METS format
- Any Export retains a full copy of the vendor-specific mark-up within the DExT-METS file
- Vendor programs should in time be capable of importing standard compliant DExT-METS. At a minimum this includes the content from the core data concepts



Technical Approach

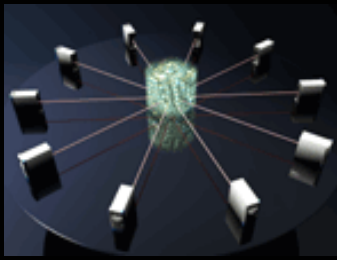
Feedback on DExT model will enable progress to be made on technical platform decisions. Considerations moving forward from the initial demonstration GUI include

- Relational or XML indexing back end (storage)
- Session-based access to studies (web enabled)
- Online access to conversion tools (client-server)
- Batch processing of studies
- Collaboration on development of tools (via SourceForge)



Planning ahead

- Looking for formal collaboration with software creators and vendors
- Further use case examples relating to the possibilities of an independent interchangeable qualitative data XML Schema
- Opensource products
- Formal implementation of the model in data archives - UKDA and we hope others to follow
- A small scale evaluation of the models and tools will be undertaken to scope out whether a functional and scalable service where data formats can be submitted and seamlessly returned in a chosen, desired format is possible



Contact

- Louise Corti
- Angat Bhat

UK Data Archive

corti@essex.ac.uk

44 1206 872145