DOES CURRENT DDI SUIT QUALITATIVE DATA?

- DDI 2 fine for describing the study and overview of a whole data collection

- good down to the individual file level (e.g. a single interview) but cannot describe the content of files, e.g. the structure of an textual interview data or how files relate to each other

- working on a descriptive standard to ensure holistic and detailed description of complex data collections

- need power to relate data, parts of data and annotations to each other

UK DATA ARCHIVE
PREVIOUS WORK ON QUALITATIVE SCHEMAS

- Data Exchange Tools (DExT) project – UK Data Archive and ODaF
  - sought to define a schema that would describe complex collections of data
    - capture relationships between data files
    - preserve references to annotations performed on data
  - purpose
    - for longer term preservation
    - for data exchange providing an open source intermediate format
AN EXAMPLE OF A COMPLEX QUALI COLLECTION

- data collection
  - 50 audio recorded interviews – 200 mp3 files
  - 50 interview transcripts – 50 word files
  - 45 summaries – 45 word files
  - 100 photos – 100 tiff files

- annotated and coded data in a CAQDAS e.g. NVivo
  - transcripts classified by some key variables
  - codes attached to segments
  - memos linked to data discussing features of the data
  - assertion - links between parts of data

- interview level metadata useful, and collected in various ways
ADDING META INFORMATION ABOUT AN INTERVIEW TO THE HEADER OF A WORD DOCUMENT

Study Name: 
Depositor: 
Interviewer: 

Interview number: 
Interview ID: Firstname Lastname 
Date of interview: 

Information about interviewee 
Date of birth: 
Gender: 
Geographic region: 

Marital status: 
Occupation: 

I: Just one or two factual details first of all before we go on to your health and that.... how old are you? 

FL: I'm 58 in June. 

I: What schools did you go to? Can you remember that far back!

UK DATA ARCHIVE
### Compiled ‘Datalist’ of Interviews in a Collection

<table>
<thead>
<tr>
<th>Interview ID</th>
<th>Gender</th>
<th>Description</th>
<th>Date of Interview</th>
<th>No of Pages</th>
<th>Text File Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>x001</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>07/09/2005</td>
<td>36</td>
<td>6124int001</td>
</tr>
<tr>
<td>x002</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>09/09/2005</td>
<td>41</td>
<td>6124int002</td>
</tr>
<tr>
<td>x003</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>09/09/2005</td>
<td>39</td>
<td>6124int003</td>
</tr>
<tr>
<td>x004</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>13/09/2005</td>
<td>36</td>
<td>6124int004</td>
</tr>
<tr>
<td>x005</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>13/09/2005</td>
<td>34</td>
<td>6124int005</td>
</tr>
<tr>
<td>x006</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>14/09/2005</td>
<td>50</td>
<td>6124int006</td>
</tr>
<tr>
<td>x007</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>21/09/2005</td>
<td>31</td>
<td>6124int007</td>
</tr>
<tr>
<td>x008</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>21/09/2005</td>
<td>35</td>
<td>6124int008</td>
</tr>
<tr>
<td>x009</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>22/09/2005</td>
<td>33</td>
<td>6124int009</td>
</tr>
<tr>
<td>x010</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>22/09/2005</td>
<td>23</td>
<td>6124int010</td>
</tr>
<tr>
<td>x011</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>22/09/2005</td>
<td>36</td>
<td>6124int011</td>
</tr>
<tr>
<td>x012</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>27/09/2005</td>
<td>41</td>
<td>6124int012</td>
</tr>
<tr>
<td>x013</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>27/09/2005</td>
<td>21</td>
<td>6124int013</td>
</tr>
<tr>
<td>x014</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>30/09/2005</td>
<td>20</td>
<td>6124int014</td>
</tr>
<tr>
<td>x015</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>05/10/2005</td>
<td>19</td>
<td>6124int015</td>
</tr>
<tr>
<td>x016</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>05/10/2005</td>
<td>27</td>
<td>6124int016</td>
</tr>
<tr>
<td>x017</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>07/10/2005</td>
<td>27</td>
<td>6124int017</td>
</tr>
<tr>
<td>x018</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>17/10/2005</td>
<td>11</td>
<td>6124int018</td>
</tr>
<tr>
<td>x019</td>
<td>Male</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>18/10/2005</td>
<td>33</td>
<td>6124int019</td>
</tr>
<tr>
<td>x020</td>
<td>Female</td>
<td>Interview with a Hospital Doctor in a Multi-Ethnic Northern City</td>
<td>07/11/2005</td>
<td>21</td>
<td>6124int020</td>
</tr>
<tr>
<td>z001</td>
<td>Male</td>
<td>Interview with Hospital Doctor in Northern Tourist City</td>
<td>19/07/2005</td>
<td>50</td>
<td>6124int021</td>
</tr>
<tr>
<td>z002</td>
<td>Male</td>
<td>Interview with Hospital Doctor in Northern Tourist City</td>
<td>10/08/2005</td>
<td>45</td>
<td>6124int022</td>
</tr>
<tr>
<td>z003</td>
<td>Male</td>
<td>Interview with Hospital Doctor in Northern Tourist City</td>
<td>17/08/2005</td>
<td>48</td>
<td>6124int023</td>
</tr>
</tbody>
</table>
Excerpt from interview transcript

My father was, in the daytime he was a boilermaker on the old North Circular Railway and then every night he played in the theatre orchestra. And sometimes after the theatre he would go on and play for an hour or two at a dance, well they called them balls in those days. And he had to be at work at six the next morning! Cornet player.

Goodness, would you see him when he used to play?

Excerpt with XML mark-up

<s n="31">My father was, in the daytime he was a boilermaker on the old North Circular Railway and then every night he played in the theatre orchestra. And sometimes after the theatre he would go on and play for an hour or two at a dance, well they called them balls in those days. And he had to be at work at six the next morning! Cornet player.

Goodness, would you see him when he used to play?</s>

<s n="44">And sometimes even after the theatre he would go on and play for an hour or two at a dance, well they called them balls in those days.
</s>

<s n="45">And he’d to go to work at six the next morning! Cornet player.</s>
Right, it starts with your grandparents. So give me the names and dates of birth of both. Do you remember those sets of grandparents?

Yes.

Well, we'll start with your mum's parents? Where did they live?

They lived in Widness, Lancashire.

How do you remember them?

When we Mum used to take me to see them and me Grandma came to live with us in the end, didn't she?

Yes, when Granddad died - 1948.

So he died when he was 48?

No, he was 52. He died in 1948.

But I remember it. How old would I be then?

Oh, you would have been little then.

I remember him, he used to have whiskers. He used to put me on his knee and give me a kiss.

Well, he'd have whiskers where he probably had not had a shave and he used to rub her cheek - literally.

What was his occupation?

He worked at ICI, Widness.

Doing what?

What was Grandmother's occupation?

She had nought, because she had arthritis. She couldn't walk. That's why I used to go home every weekend to look after her. We had a Home Help until Friday and I used to go Saturday and stop until Monday when the Home Help came again.

Really? This was when you were married?

Yes. I took me children with me. Me husband, give him his due, he let me go every Saturday, tea-time, until Monday and I used to take the children with me.

And then she came to live with you eventually?

Well, me Dad died. Me sister was younger than me and she got TB and she went in Rafford Sanatorium, but I still come every week and looked after me Mum on a Saturday night. Got up, got the dinner ready and got the bus from Widness to Rafford every Sunday to see me sister. We, she was in for two years in Rafford and then they said they couldn't do nothing, so I brought her home and I nursed her for six weeks and she died and that's when we Mum came to me and we give the house up. There was nothing left. I got three brothers, but wifes are not the same as their own daughter. So we all fell out over who was going to have her. So I said well, nobody's having her only me. I brought her to Altrincham and she lived two years. I made her comfortable in the kitchen with a single bed and looked after her. Gave me little job up and I looked after her and she just lived two years.
CURRENT EXCHANGEABILITY IN QUAL SOFTWARE

• CAQDAS software are being used as data management systems for projects
• import files into system and they gets ‘locked’ into its proprietary databases
• Atlas.ti and Nvivo export XML, though not using an agreed schema
• ATLAS.ti first vendor to pioneer data exchange by exporting annotations in XML (MUHR, 2000)

• Need data exchange standard to manage complex data
QuDex SCHEMA

- QuDex schema V3 published in 2006
- core features
- various refinements
- basic viewer available
- Arofan will go through…
DDI QUALITATIVE WORKING GROUP

- Set up in late 2009
- first meeting April 2010 via skype
- 21 members across 17 locations and 9 countries
- collected use cases for complex qualitative collections from group members
- mapped use cases to the DDI Lifecycle stage elements (thanks to Larry)
USE CASES REFERENCING LIFE CYCLE ELEMENTS
# QuDex ELEMENTS AND DEFINITIONS

<table>
<thead>
<tr>
<th>Top level Elements</th>
<th>Sub elements</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>&lt;qudex&gt;</code></td>
<td>resourceCollection, segmentCollection, codeCollection, memoCollection, categoryCollection, relationCollection</td>
<td>The root element; a ‘wrapper’ for all other elements of the QuDEx Schema. Each top level element in QuDEx is defined as a ‘collection’ and must appear in the order outlined below</td>
</tr>
<tr>
<td><code>&lt;resourceCollection&gt;</code></td>
<td>sources, memoSources, documents</td>
<td>The <code>resourceCollection</code> section lists and locates all content available to the QuDEx file. A <code>source</code> points to the original location of the resource while each author working on the QuDEx file is assigned a surrogate <code>document</code> which points to the relevant <code>source</code>. The child elements <code>sources</code> and <code>memoSources</code> contain direct references to the files under analysis; the <code>documents</code> section contains their surrogates</td>
</tr>
<tr>
<td><code>&lt;segmentCollection&gt;</code></td>
<td>Segment (sub elements text, audio, video, xml, image)</td>
<td>The parent element for all segments, which is a subset of a document (text, audio, video or image) under analysis defined in a manner appropriate to the format (text, audio, video, image or xml). Segments may overlap and multiple memos and codes may be assigned to a segment. Start and end points can be formally assigned to segments of text, and audio visual materials in other document</td>
</tr>
<tr>
<td><code>&lt;codeCollection&gt;</code></td>
<td>code</td>
<td>The parent element for all codes. A code is a short alphanumeric string, usually a single word; may be assigned to a segment or document though assignment is not required. A code may optionally be taken from a controlled vocabulary defined under @ authority</td>
</tr>
</tbody>
</table>
## QuDex ELEMENTS AND DEFINITIONS

<table>
<thead>
<tr>
<th>Element Collection</th>
<th>Element/Description</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;memoCollection&gt;</td>
<td>mem (sub elements memoDocumentRef, memoText)</td>
<td>The parent element for all <em>memos</em>; these may be pure text and embedded in the QuDEx file (inline memo) or may refer to external files. A memo is a text string internal to the document (inline memo) or an externally held document (external memo) which may be assigned to a segment, code, document, category or to another.</td>
</tr>
<tr>
<td>&lt;categoryCollection&gt;</td>
<td>category</td>
<td>The parent element for all categories. A category is an alphanumeric string (stored in @label) assigned to one or more documents. Categories may be hierarchically nested. Documents contained within a category are referenced using @documentRefs. Nested categories are referenced using @categoryRefs.</td>
</tr>
</tbody>
</table>
| <relationCollection>   | objectRelation                                                                      | The parent element for all relationships between objects. For the purposes of a relation all of the following are considered to be ‘objects’:  
  - A *document*: surrogate of a *source* or *memoSource*  
  - A *segment* within a *document*  
  - An assigned value: *code*, *memo*, *category*, *relation*  
  A relation is a link between two objects in a QuDEx file. Each object is either the start or end point of a relation (source vs target). Every relation may, optionally, have a name. |
WORKING WITH QuDex

- taking QuDex as starting point as it is useful for both whole document and ‘part of document’ description
- some members worried about complexity of QuDex, but reassured that it can do very basic file level description
- demonstrator work by some members on DDI-QuDex based FEDORA ingest systems
FEDORA INGEST WORK

Two current UK data archiving projects using DDI and QuDex metadata in FEDORA

- Ensemble project, Liverpool. Martinez and Gregory
  - QuDex Repository: FEDORA based framework

- Timescapes archive, Leeds. Ben Ryan
  - Conversion of existing Digitool qualitative data archive to FEDORA
ENSEMBLE QuDex REPOSITORY BASIC FUNCTIONS

• **ingest** of collections of files from a collection along with associated metadata at study and file level

• **load/transform/index** the data and metadata ingested, making it available as a set of objects in FEDORA repository, and exposing it for use as RDF

• **dissemination to the end user** through various tools
  • search repository for studies and files
  • locate and use contents in various applications
PILOT DELIVERABLES

- spreadsheet-based tool for capturing metadata about qualitative studies and files. Used for ingest into the FEDORA repository

- metadata display tool using Exhibit browser (a Simile widget for maps, timelines and faceted browsing)

- tool for harvesting study level DDI 1/2 and DC metadata from XML instances (qual and quant)

- interface for web-based searches through the repository, designed to be integrated into own websites. Uses Lucene

- automatically populated and managed Mulgara triple-store
  - mirroring the contents of the Repository
  - exposes the contents as RDF in a SPARQL end-point and as Exhibit compatible JSON
OTHER METADATA SCHEMA

- DDI3, of course!

- TEI for annotated texts. New P5 version uses similar stand off mark-up notation

- Open Annotation Consortium standards, for annotating video

- FOXML and METS in FEDORA work
OTHER METADATA CREATION USING DDI/QuDEx

• The UK Data Archive’s Qualidata Publisher
  • UK Data Archive has built and are testing an open-source Flex-based system for processing textual interview data
  • uses DDI2 study level data and TEI document level elements
  • could easily use QuDex elements
  • produces user-ready format (.rtf) and XML versions of transcripts and documents item level redaction and access restrictions
  • Please try it out – it looks cool build in Adobe Flex
NEXT STEPS FOR THE DDI GROUP

• evaluate DDI3 definitions to see which physical instances might be relevant

• select most popular use cases to focus on

• translate into technical use cases

• bring to TIC; choose most appropriate tools development based on most suitable metadata

• Special Quali DDI focused workshop in December in Gotheburg

• your input!