Integrating DDI with metadata for qualitative data: implementing the QuDEx standard

Louise Corti, UK Data Archive
Agustina Martinez, Liverpool John Moores
Daren Bell, UK Data Archive

EDDI13, Paris
December 2013

UK Data Service

Digital futures project: rationale & aims

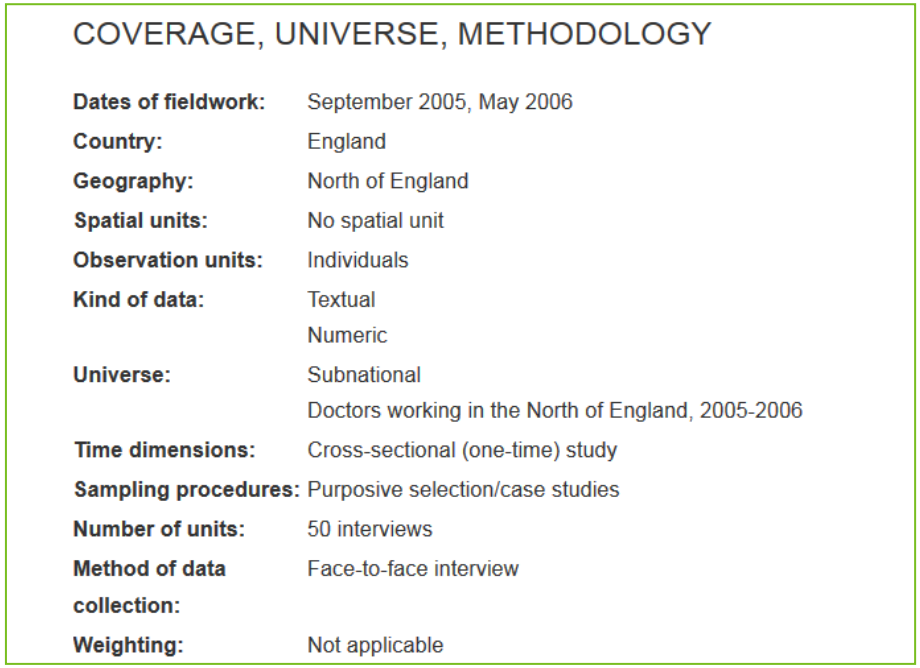
- 18 month wham-bang capital project
- Provide enhanced access to key qualitative data via online data browsing and exploration: UK Quali Bank
- Based on existing metadata schemas, known technologies. Oh, and expert consultants
- Offer a mechanism for reliably citing data located in the system
- Project includes large-scale digitisation of precious and undigitized materials
- Maximise the impact from existing research and resource investments – demonstrate re-use

UK Data Service: qualitative data services

- We have one of the largest qualitative data collections - over 350 data collections
- Currently users find and download these from our website
  - not so easy to find, but study documentation good
  - no searching within collections
  - no file manifest shown until download
  - can be a bit of guess work!
- Have Datacite collection DOIs; cannot cite parts of data

Finding & accessing qualitative data

- Search for “health” in our data catalogue, Discover
- Retrieve a study-level catalogue record
- DDI 2.x very limited for describing file content: no var lists



COVERAGE, UNIVERSE, METHODOLOGY

Dates of fieldwork:	September 2005, May 2006
Country:	England
Geography:	North of England
Spatial units:	No spatial unit
Observation units:	Individuals
Kind of data:	Textual
	Numeric
Universe:	Subnational
	Doctors working in the North of England, 2005-2006
Time dimensions:	Cross-sectional (one-time) study
Sampling procedures:	Purposive selection/case studies
Number of units:	50 interviews
Method of data collection:	Face-to-face interview
Weighting:	Not applicable

- View limited data list and user guide
- Web download as RTF bundle (46 transcripts)

Data listing

Study Number 6124
Being a Doctor: a Sociological Analysis, 2005-2006
Nettleton, S

Interview ID	Gender	Description	Date of Interview	No of Pages	Text File Name
x001	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/09/2005	36	6124int001
x002	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	41	6124int002
x003	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	09/09/2005	39	6124int003
x004	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	36	6124int004
x005	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	13/09/2005	34	6124int005
x006	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	14/09/2005	50	6124int006
x007	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	31	6124int007
x008	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	21/09/2005	35	6124int008
x009	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	33	6124int009
x010	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	23	6124int010
x011	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	22/09/2005	36	6124int011
x012	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	41	6124int012
x013	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	27/09/2005	21	6124int013
x014	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	30/09/2005	20	6124int014
x015	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	19	6124int015
x016	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	05/10/2005	27	6124int016
x017	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/10/2005	27	6124int017
x018	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	17/10/2005	11	6124int018
x019	Male	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	19/10/2005	33	6124int019
x020	Female	Interview with a Hospital Doctor in a Multi-Ethnic Northern City	07/11/2005	21	6124int020
z001	Male	Interview with Hospital Doctor in Northern Tourist City	19/07/2005	50	6124int021
z002	Male	Interview with Hospital Doctor in Northern Tourist City	10/08/2005	45	6124int022
z003	Male	Interview with Hospital Doctor in Northern Tourist City	17/08/2005	56	6124int023
z004	Male	Interview with Hospital Doctor in Northern Tourist City	07/11/2005	21	6124int024
z005	Female	Interview with Hospital Doctor in Northern Tourist City	14/11/2005	32	6124int025
z006	Male	Interview with Hospital Doctor in Northern Tourist City	15/11/2005	23	6124int026
z007	Male	Interview with Hospital Doctor in Northern Tourist City	16/11/2005	23	6124int027
z008	Female	Interview with Hospital Doctor in Northern Tourist City	17/11/2005	18	6124int028
z009	Male	Interview with Hospital Doctor in Northern Tourist City	18/11/2005	20	6124int029

Complex data collections

- SN 5801: Concepts of Healthy Eating Food Research: Phases I and II, 1992-1996
- 293 interview transcripts; 73 diaries; 6 field notes
- Not represented well at all in a DDI 2.X catalogue

Geography: Lewisham; Newport; Greater London; Pembrokeshire
Spatial Units: Unitary Authorities (England);Unitary Authorities (Wales)
Observation Units: Individuals; Families/households
Kind of Data: Textual data; Individual (micro) level; open-ended, semi-structured interviews; diaries

Universe Sampled:
*Location of Units of Observation:*Subnational
*Population:*Men and women of all ages, black or white British, English and/or Welsh speaking, middle-and working-class backgrounds; Retail, catering and health professionals; Tourists in the Newport area.

Methodology:
Time Dimensions: Cross-sectional (one-time) study
Sampling Procedures: Convenience sample
Number of Units: 293 interview transcripts; 73 diaries; 6 observation field notes
Method of Data Collection: Face-to-face interview; Observation; Diaries
Weighting: No weighting used

Download Zip of data and doc

Format download	Filename	File size
RTF	6124RTF_19F5C62AEC1EDF34B42B4156FAF35760.ZIP	2.10 MB

Get Adobe Reader Get Adobe Acrobat Reader

Get StuffIt Expander

New Open Add Extract Test Exit

4943RTF_635968BB383BDA0CDE
- UKDA-4943-rtf
  - mrdoc
    - excel
    - pdf
    - UKDA
  - rtf

UKDA-4943-rtf\

File Name	File Type	Modified	Size	Ratio	Packed	CRC	Attributes	Folder
mrdoc	Folder						D	UKDA-49...
rtf	Folder						D	UKDA-49...
4943_file_information.rtf	Rich Text Format	30/04/2008 12:...	15,931	90%	1,578	3CAC6FC6		UKDA-49...
read4943.txt	Text Document	30/04/2008 11:...	7,348	60%	2,966	EA120EDE		UKDA-49...

UKDA-6124-rtf\rtf\

File Name	File Type	Modified	Size	Ratio	Packed	CRC	Folder
6124int050.rtf	Rich Text Format	11/03/2009 12:11...	178,651	83%	29,838	F8FEE4C3	UKDA-61...
6124int049.rtf	Rich Text Format	11/03/2009 12:11...	135,117	83%	22,786	A85A7C75	UKDA-61...
6124int048.rtf	Rich Text Format	11/03/2009 12:11...	157,114	83%	27,340	D7B142EB	UKDA-61...
6124int047.rtf	Rich Text Format	11/03/2009 12:10...	145,465	84%	22,728	6E2E3D04	UKDA-61...
6124int046.rtf	Rich Text Format	11/03/2009 12:10...	130,052	83%	22,722	7687F028	UKDA-61...
6124int045.rtf	Rich Text Format	11/03/2009 12:10...	246,005	89%	27,341	5C8FF225	UKDA-61...
6124int044.rtf	Rich Text Format	11/03/2009 12:10...	346,138	89%	39,173	203CF033	UKDA-61...
6124int043.rtf	Rich Text Format	11/03/2009 12:09...	272,140	89%	28,957	ADFF70B9	UKDA-61...
6124int042.rtf	Rich Text Format	11/03/2009 12:09...	268,071	89%	29,985	C04A51CF	UKDA-61...
6124int041.rtf	Rich Text Format	11/03/2009 12:09...	185,379	89%	19,919	832C0D48	UKDA-61...
6124int040.rtf	Rich Text Format	11/03/2009 12:08...	288,020	90%	28,088	C27DEF0E	UKDA-61...
6124int039.rtf	Rich Text Format	11/03/2009 12:08...	229,536	89%	26,060	595569FF	UKDA-61...
6124int038.rtf	Rich Text Format	11/03/2009 12:08...	194,189	89%	22,188	F0EF3B80	UKDA-61...
6124int037.rtf	Rich Text Format	11/03/2009 12:08...	262,463	89%	28,930	6FF37118	UKDA-61...
6124int036.rtf	Rich Text Format	11/03/2009 12:07...	221,079	89%	24,497	803993DB	UKDA-61...
6124int035.rtf	Rich Text Format	11/03/2009 12:07...	192,414	89%	20,384	774AF888	UKDA-61...
6124int034.rtf	Rich Text Format	11/03/2009 12:07...	286,909	90%	29,868	B28CA117	UKDA-61...
6124int033.rtf	Rich Text Format	11/03/2009 12:07...	277,765	89%	30,427	DCD54562	UKDA-61...
6124int032.rtf	Rich Text Format	11/03/2009 12:06...	209,945	90%	21,700	410EAC13	UKDA-61...
6124int031.rtf	Rich Text Format	11/03/2009 12:06...	153,461	88%	18,116	140E55DC	UKDA-61...
6124int030.rtf	Rich Text Format	11/03/2009 12:06...	253,723	89%	26,692	653178D2	UKDA-61...

Study Name:
Depositor:
Interviewer:

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

Information about interviewee
Date of birth:
Gender:
Marital status:
Occupation:
Geographic region:

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!

FL: Oh... the last school was at Longside, aye, ken Longside?

I: No, where is that?

FL: Peterheid Village... That was the last school.

I: Uh-huh, so you lived in Peterhead...

FL: No, Longside.

I: Longside. And, do you work at all? At the moment?

FL: Just look efter my grandchildren. Like that... well, my grand-daughter comes in at night, well, her mither cleans the school... and I look efter my grandson whose mother works... aye, that kind of thing. I take him in an' keep an eye on him. Well, he's at the school but I give him his dinner an' look efter him at night till she comes an' picks him up.

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!

FL: Oh... the last school was at Longside, aye, ken Longside?

I: No, where is that?

FL: Peterheid Village... That was the last school.

Digitisation of key data sources

Selectively digitize paper-based materials:

- original survey questionnaires; open ended questions
- transcribed interviews
- handwritten field notes, essays
- diagrams
- photographs

Destination formats:

- all text files: XML
- handwriting/photos/pictures: images as jpeg
- multipage non-XMLable docs: images as PDF
- audio: mp3

Metadata demands for UK Quali Bank

Discover, locate, retrieve, examine in context, cite

- Requires rich descriptive metadata for files
- Ability to display and link data objects: text to related audio, images, and other research outputs
- Highly structured and consistently marked-up data
- Use of common metadata elements enable federated catalogues across providers and borders

Description below the collection

- DDI Codebook 2.5 for basic study-level catalogue metadata
- QuDEx schema (Qualitative Data Exchange) for rich file-level description, document coding and annotation and intra-collection relationships
- Limited use of TEI schema (Text Encoding Initiative) for mark-up of textual data items

QuDEx Overview

- XML schema for documenting metadata for qualitative data collections
  - developed as part of the JISC 2007 Data Exchange Tools (DExT) project – UK Data Archive and MetadataTechnologies
- QuDEx is intended to be used as a standard way of encoding metadata for qualitative collections
  - for exchange between CAQDAS packages
  - for use within data archives and libraries
  - for dissemination systems
- Allow for describing complex collections of quali data
  - detailed description at the object level, e.g. interview characteristics, interview setting, type of object etc.
  - capture relationships between resources (files)
  - preserve references to annotations performed on data

QuDEx Structure

- A qualitative study can be described using a single QuDEx instance (the XML document)
  - <resourceCollection>
  - <segmentCollection>
  - <codeCollection>
  - <categoryCollection>
  - <memoCollection>
  - <relationCollection>

Schema at: www.data-archive.ac.uk/create-manage/projects/qudex/

QuDEx – Resource Collection

- Resource collection is broken into two sections, sources and documents:
  - a <source> describes a physical file of the qualitative data
  - a <document> is a file available to an application. It is typically a copy of a source, which might have been re-formatted, and it always references its <source>
- Each data file included in a data collection has to be described as a pair of <source> and <document> within a study's QuDEx instance

QuDEx – Category Collection

- A <category> is a formal code applied to the qualitative data (usually documents or segments) by the researcher
  - usually taken from a formal classification or controlled vocabulary
  - can be applied to documents, segments, codes, memos, or other categories using a <relation>
  - can be used to group documents directly, by simply listing them
  - can be arranged in hierarchies

QuDEx – Memo Collection

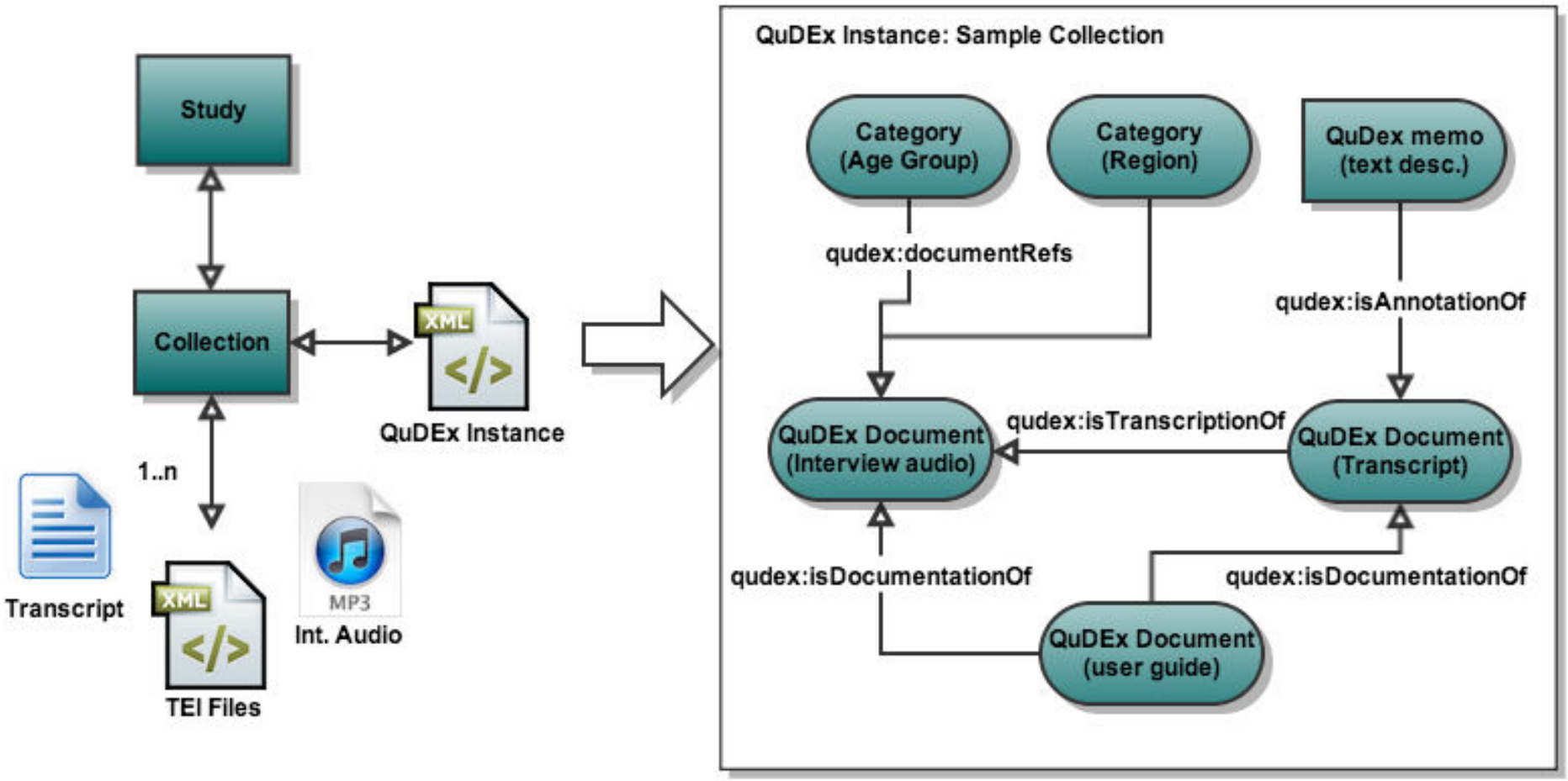
- A <memo> is used to store annotations of data files, e/g those produced by the researcher.
  - examples include: analytic notes, interview notes, or any richer descriptive information
  - can be held in the QuDEx document (inline) or stored in an external file (memo source)
  - can be associated with documents, segments, memos, categories, or codes using a <relation> element

QuDEx – Relation Collection

- A <relation> element links a source and target, using a typed relationship
  - there is a fixed list of typical relationships (i.e. “isPartOf”, “isFormatOf”)
  - user-defined relationships are also allowed (i.e. by using “otherRelationName” element). E.g. “isTranscriptionOf”, “isAnnotationOf”…
  - optionally, the type of the relationship can be specified. E.g. “documentDocument”, “categoryDocument”, or “memoDocument”
  - source and target may be documents, segments, categories, memos, or codes

Use of QuDEx

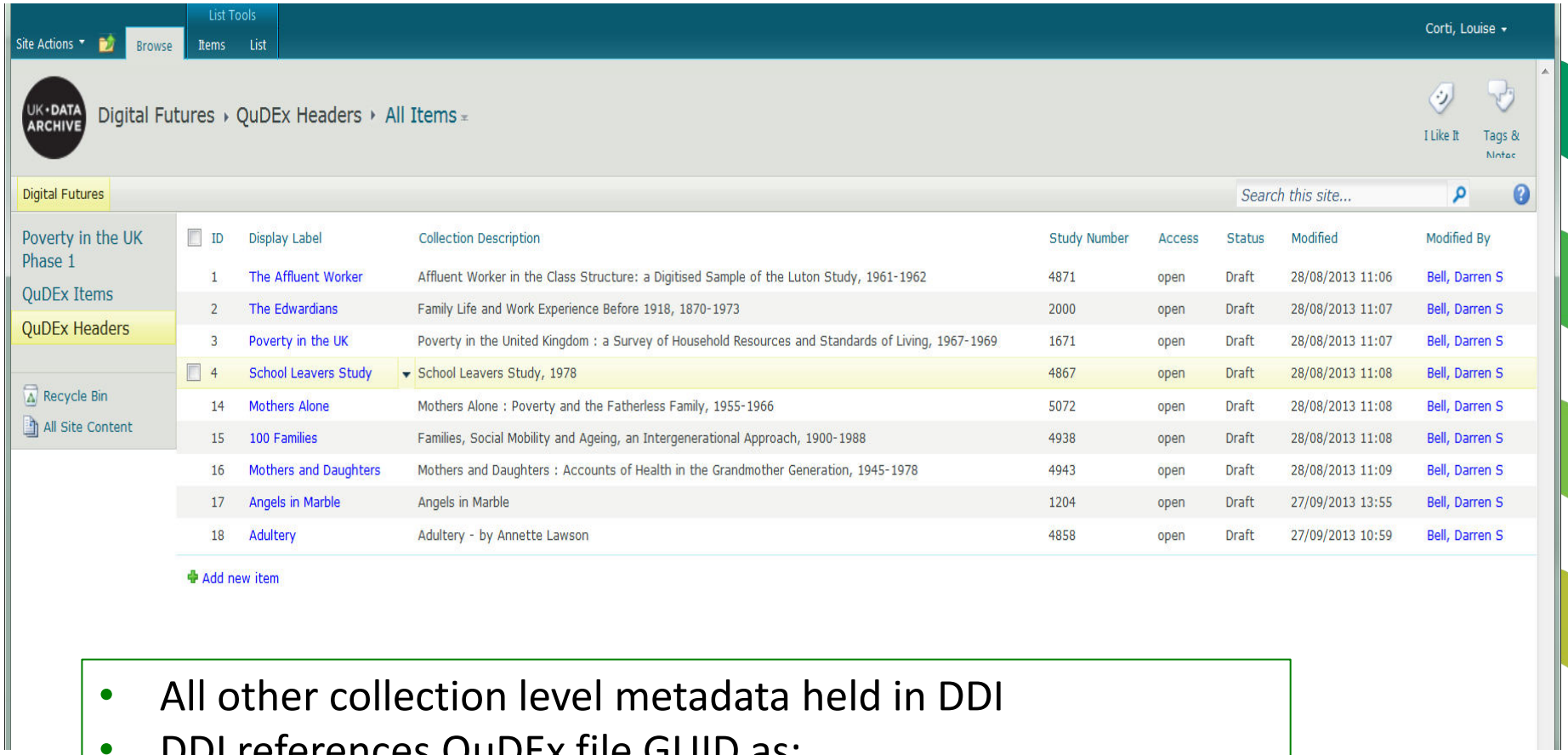
- QuDEx instance is central to the Quali Bank system. It provides:
  - access to study-level documentation by referencing existing DDI 2.5 documents
  - data listing for resources included in a collection (by using <sources> and <documents> elements)
  - attachment of categories to data resources and collection documentation files
  - specification of data relationships (by using <relation> elements)


Study
Collection
XML
QuDEx Instance
1..n
Transcript
XML
TEI Files
MP3
Int. Audio
QuDEx Instance: Sample Collection
Category (Age Group)
Category (Region)
QuDex memo (text desc.)
qudex:documentRefs
qudex:isAnnotationOf
QuDEx Document (Interview audio)
qudex:isTranscriptionOf
QuDEx Document (Transcript)
qudex:isDocumentationOf
qudex:isDocumentationOf
QuDEx Document (user guide)

System expectations

- Search/browse for data
  - Search /faceted browse of data - text; image/PDF, audio
- Browse
  - Faceted browse by:
    - Collection level: title, date and openness
    - Object level: data type, interview characteristics, location
- Search:
  - Display no. hits and minimal item metadata
  - Word in paragraph; thumbnail image/pdf; AV link
  - Context: other related objects, within system or external
- Access full object
  - View data, key metadata and all related files and links
  - Link to download full collection and user documentation (UKDS)
  - Get citation for part of data

QuDEx collection level metadata



- All other collection level metadata held in DDI
- DDI references QuDEx file GUID as:

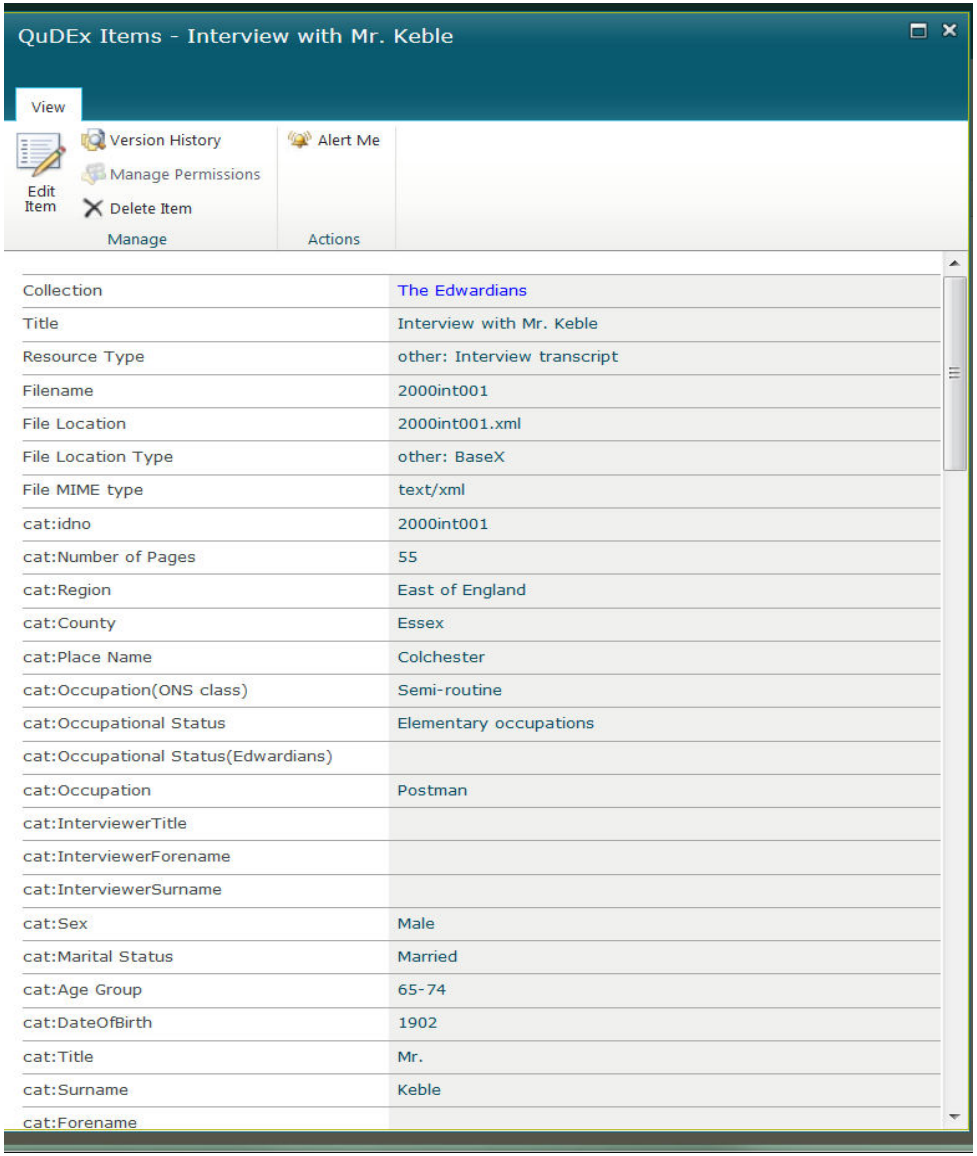
```
<othrStdyMat>
    <relmat ype = “qudex”>GUID</relmat>
```

UK Data Service

Within collection object metadata

Collection	Title	Resource Type	Filename	File Location	File Location Type	File MIME type	cat:idno	cat:Number of Pages	cat:Region	cat:County	cat:Place Name	cat:Occupation(ONS class)	cat:Occupation
The Edwardians	Interview with Mr. Keble	other: Interview transcript	2000int001	2000int001.xml	other: BaseX	text/xml	2000int001	55	East of England	Essex	Colchester	Semi-routine	Elementary occ
The Edwardians	Interview with Mr. Stinchcombe	other: Interview transcript	2000int002	2000int002.xml	other: BaseX	text/xml	2000int002	109				Semi-routine	Elementary occ
The Edwardians	Interview with Mrs. Duckers	other: Interview transcript	2000int003	2000int003.xml	other: BaseX	text/xml	2000int003	78	London	Greater London area	London	Routine	Elementary occ
The Edwardians	Interview with Mr. Knifton	other: Interview transcript	2000int004	2000int004.xml	other: BaseX	text/xml	2000int004	74	London	Greater London area	London	Routine	Skilled trades occupations
The Edwardians	Interview with Mr. Troy	other: Interview transcript	2000int005	2000int005.xml	other: BaseX	text/xml	2000int005	85	London	Greater London area	London	Semi-routine	Elementary occ
The Edwardians	Interview with Mrs. Craig	other: Interview transcript	2000int006	2000int006.xml	other: BaseX	text/xml	2000int006	61	London	Greater London area	London	Semi-routine	Skilled trades occupations
The Edwardians	Interview with Mrs. Wilson	other: Interview transcript	2000int007	2000int007.xml	other: BaseX	text/xml	2000int007	73	London	Greater London area	London	Routine	Elementary occ
The Edwardians	Interview with Mr. Rook	other: Interview transcript	2000int008	2000int008.xml	other: BaseX	text/xml	2000int008	82	London	Greater London area	London	Routine	Elementary occ
School Leavers Study	Image relating to Essay by 1	image	4867ess001	4867ess001.jpg	url	image/jpeg	4867ess001		South East	Kent	Isle of Sheppey	Not classified	Student
School Leavers Study	Essay by 1	other: Essay	4867ess001	4867ess001.xml	other: BaseX	text/xml	4867ess001		South East	Kent	Isle of Sheppey	Not classified	Student
School Leavers Study	Image relating to Essay by 2	image	4867ess002	4867ess002.jpg	url	image/jpeg	4867ess002		South East	Kent	Isle of Sheppey	Not classified	Student
School Leavers Study	Essay by 2	other: Essay	4867ess002	4867ess002.xml	other: BaseX	text/xml	4867ess002		South East	Kent	Isle of Sheppey	Not classified	Student
School	Image	image	4867ess003	4867ess003.jpg	url	image/jpeg	4867ess003		South	Kent	Isle of	Not classified	Student

QuDEx Category schema



Collection	The Edwardians
Title	Interview with Mr. Keble
Resource Type	other: Interview transcript
Filename	2000int001
File Location	2000int001.xml
File Location Type	other: BaseX
File MIME type	text/xml
cat:idno	2000int001
cat:Number of Pages	55
cat:Region	East of England
cat:County	Essex
cat:Place Name	Colchester
cat:Occupation(ONS class)	Semi-routine
cat:Occupational Status	Elementary occupations
cat:Occupational Status(Edwardians)	
cat:Occupation	Postman
cat:InterviewerTitle	
cat:InterviewerForename	
cat:InterviewerSurname	
cat:Sex	Male
cat:Marital Status	Married
cat:Age Group	65-74
cat:DateOfBirth	1902
cat:Title	Mr.
cat:Surname	Keble
cat:Forename	

- Available for all collections
- Primary core metadata.
- Can add to for additional (secondary) metadata
- Metadata population - a lot of manual work!
- Little that can be automated unless from a linked quant file
- Object relations. AutoGenerates converse relation

isTranscriptOf - hasTranscription

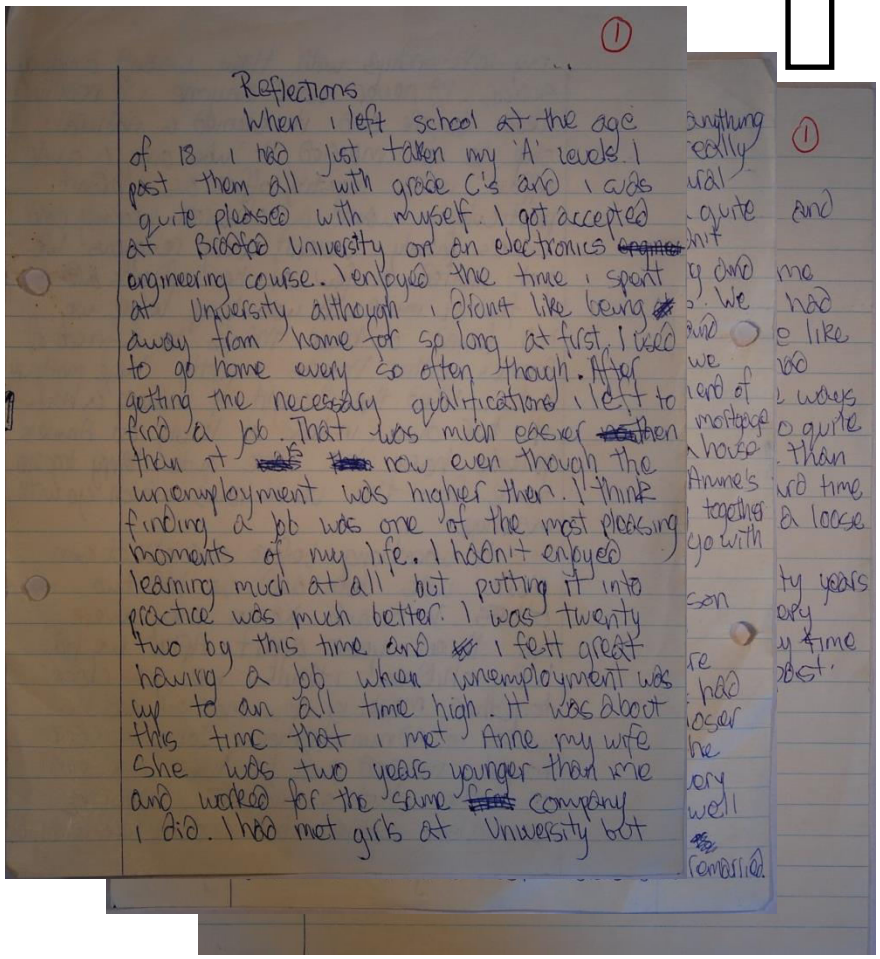
Use of Text Encoding Initiative (TEI)

- Bespoke TEI schema. Minimal use of TEI tags, from a massive profile

- To denote structural mark-up
  - TEI header: 3 mandatory elements
  - Body elements: Turn takers, paragraphs, headers
  - Inline tags: Corrections, errors

- Use of unique GUIDs to identify both TEI and QuDEx objects:
  - Collection, Files, Paragraphs

School Leavers on the Isle of Sheppey



Reflections

When I left school at the age of 18 I had just taken my 'A' levels. I past them all with grade C's and I was quite pleased with myself. I got accepted at Bradford University on an electronics engineering course. I enjoyed the time I spent at University although I didn't like being away from home for so long at first. I used to go home every so often though. After getting the necessary qualifications I left to find a job. That was much easier than it is now even though the unemployment was higher then. I think finding a job was one of the most pleasing moments of my life. I hadn't enjoyed learning much at all but putting it into practice was much better. I was twenty two by this time and I felt great having a job when unemployment was up to an all time high. It was about this time that I met Anne my wife. She was two years younger than me and worked for the same company I did. I had met girls at University but

Name of project: School Leavers Study, 1978
Depositor: Pahl, R.E.
ID: Essay 1
Group: Boys
Filename: 4867ess001

ttt Reflections

aaa When I left school at the age of 18 I had just taken my 'A' levels. I past them all with grade C's and I was quite pleased with myself. I got accepted at Bradford University on an electronics engineering course. I enjoyed the time I spent at university although I didn't like being away from home for so long at first. I used to go home every so often though. After getting the necessary qualifications I left to find a job. That was much easier than it is now even though the unemployment was higher then. I think finding a job was one of the most pleasing moments of my life. I hadn't enjoyed learning much at all but putting it into practice was much better. I was twenty two by this time and I felt great having a job when unemployment was up to an all time high. It was about this time that I met Anne my wife. She was two years younger than me and worked for the same company as I did. I had met girls at university but my relationships with them weren't anything serious. I'd never none anyone I'd really liked before and it seemed a natural step to get married. It was only a quite affair in a registry office as I didn't particularly want a church wedding and Anne's family weren't very religious. We just invited a few of her friends and a few of mine. It was long before we had found a house thanks to a friend of Anne's father. It was a pretty large mortgage but I liked the thought of having a house of our own, instead of living at Anne's parent's house. We were quite happy together aside from the usual rows which go with marrige.

aaa It wasn't long before our eldest son Mark was born and we had two others John and Karen, they were all three years apart. After we had had children I think we grew closer together. Mark is now married and he works in Administration. John is very happily married and he has done well in the plastics industry. Karen was also married but divorced and remarried- she is now a housewife and they have two children.

aaa It was a great experience for me to see my children growing up. I had often wondered what it would be like when I was young. When they had all left and gone their separate ways it felt very empty and it was so quite I think Anne missed them more than me even though she had had a hard time bringing them up, she seemed at a loose end.

aaa I retired not long ago after forty years with the same company. Life is very tedious now and I spend most of my time just sitting around thinking of the past.

EOF

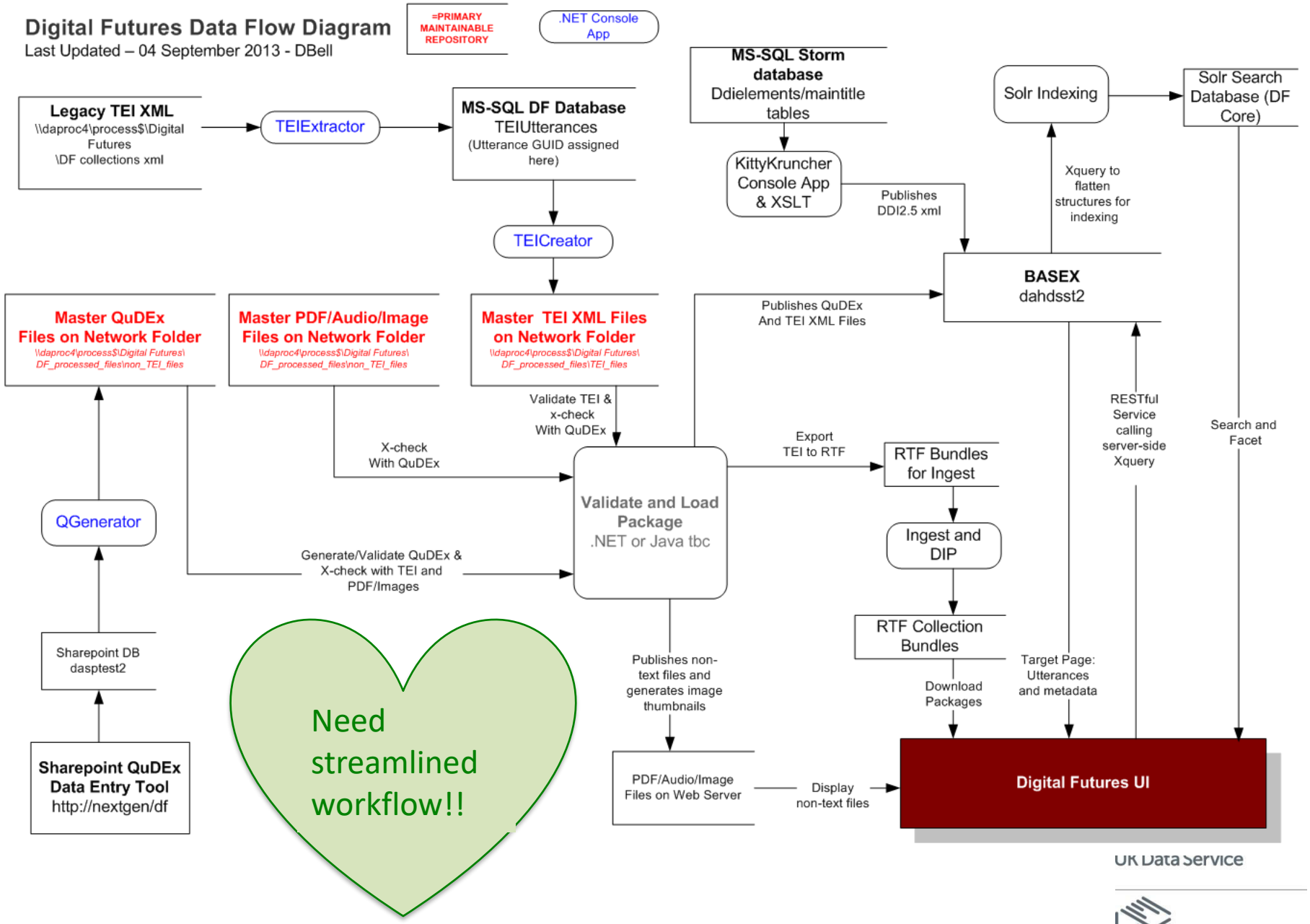
UK Data Service

XML: School Leavers on the Isle of Sheppey

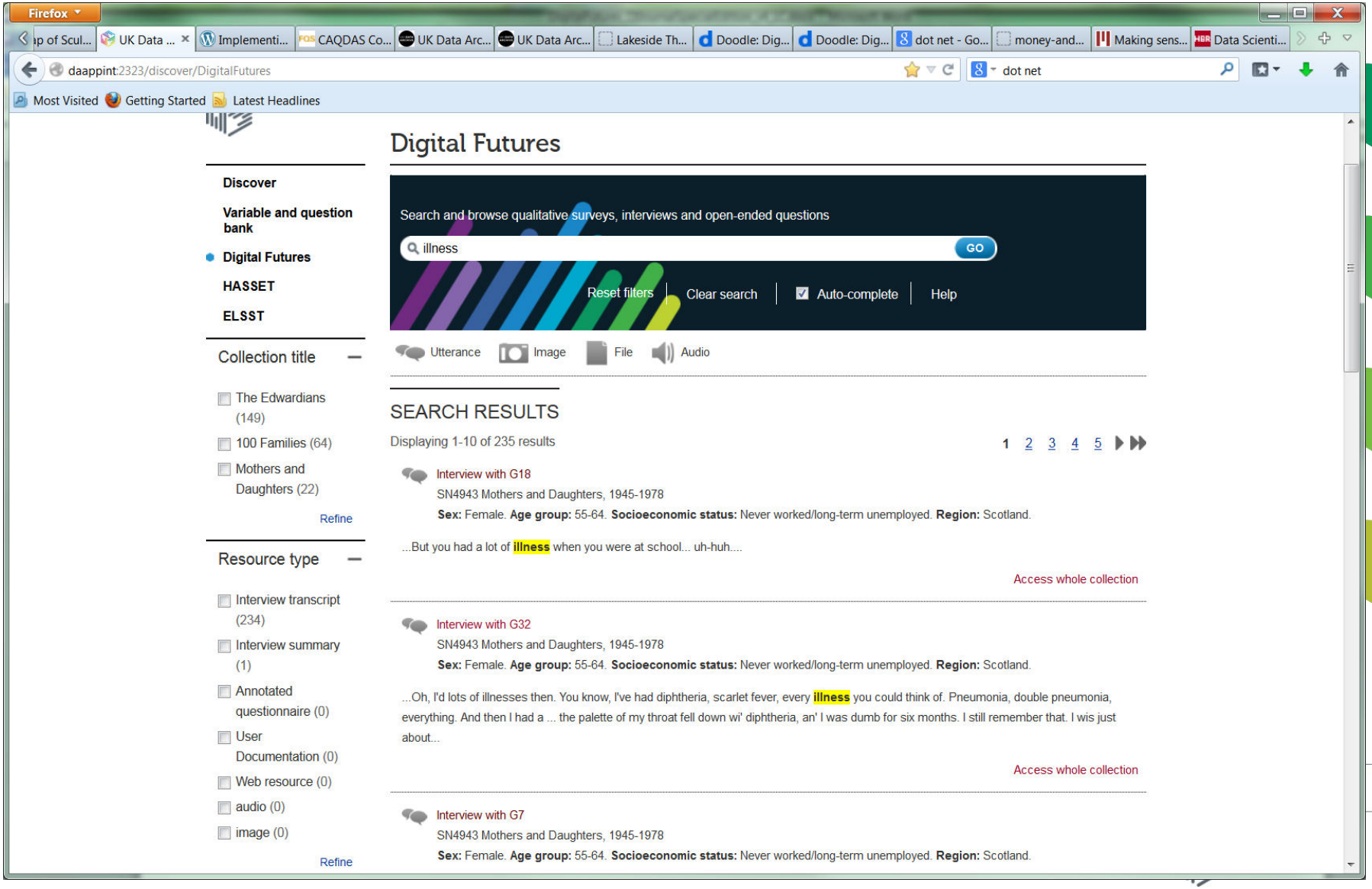
```
1  l version="1.0" encoding="utf-8"?>
2  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
3  xsi:schemaLocation="http://www.tei-c.org/ns/1.0 http://www.tei-c.org/release/xml/tei/custom/schema/xsd/tei_all.xsd"
4  xmlns="http://www.tei-c.org/ns/1.0">
5  <teiHeader>
6      <fileDesc>
7          <titleStmt>
8              <title type="collectiondisplaylabel">School Leavers Study</title>
9              <title type="sourcedisplaylabel">Essay by 5</title>
10         </titleStmt>
11         <publicationStmt>
12             <idno type="qudexcollection">q-41037361-714f-493f-94bd-18ee7e3fa29e</idno>
13             <idno type="qudex ource">q-320b1db3-e53a-436a-9d80-f050aad74db5</idno>
14             <idno type="qudex cument">q-a8754b06-46fe-40eb-be2f-b57ef8e6ded2</idno>
15         </publicationStmt>
16         <sourceDesc>
17             <p>School Leavers Study, 1978</p>
18         </sourceDesc>
19     </fileDesc>
20 </teiHeader>
21 <text>
22     <body>
23         <head>Reflections</head>
24         <u xml id="q-6c7d5d94-362c-e311-93c1-000bdb5cc6d5" who="subject" n="1">After taking my
25             Os I went on to stay in the sixth form where I took A <sic>leavle</sic><corr>level</corr> maths, physics and
26             history I <sic>narowly</sic><corr>narrowly</corr> scraped passes and by some <sic>miracel</sic><corr>miracle</corr> gaine
27             so far in my life this had been the more prolonged stay away from home but I wasn't
28             to disturbed at university I tended only to associated with a few people who I had a
29             lot in common with this is how my friendships have been for most of my life but I
30             would blame that on being an only child. <sic>Scince</sic><corr>Since</corr> a very early age I had been
31             <sic>fasenated</sic><corr>fascinated</corr> with the past and <sic>scince</sic><corr>since</corr> the age of fourteen I h
32             <sic>archaeolical</sic><corr>archaeological</corr> digs and so it followed that I got a job as <sic>asistent</sic><corr>a
33             pay wasn't much but I was <sic>realy</sic><corr>really</corr> happy the odd girlfriend came and went but as I
34             spent a great deal of time in a tent or <sic>diging</sic><corr>digging</corr> holes I <sic>didnt</sic><corr>didn't</corr>
35             one day I made the mistake and decided to settle and have a home so I tried to get a
36             post somewhere where I could continue poking in things that have <sic>hapend</sic><corr>happened</corr> but there
37             were no opening I then I felt I had wasted my education on learning something I
38             liked when I should have <sic>lernt</sic><corr>learnt</corr> <sic>somthing</sic><corr>something</corr> to make money so I
39             a probably <sic>batcheler</sic><corr>bachelor</corr> doing what I liked or as a someone doing a <sic>meanial</sic><corr>m
40             with <sic>securety</sic><corr>security</corr> and a <sic>permanate</sic><corr>permanent</corr> home on one side was the t
41             and on the other a high chance of not having a family I chose to stay as I was a
42             hoop and if the <sic>opertunity</sic><corr>opportunity</corr> came to have a job I liked where I could still go out and
43             dig but still have a <sic>permenate</sic><corr>permanent</corr> base I would take it and so when my chance came to
44             take a job in a <sic>museam</sic><corr>museum</corr> came I <sic>lept</sic><corr>leapt</corr> at it and apart for changin
45             twice I stayed that way till I retired last month. I did get married to someone who
```

System tools

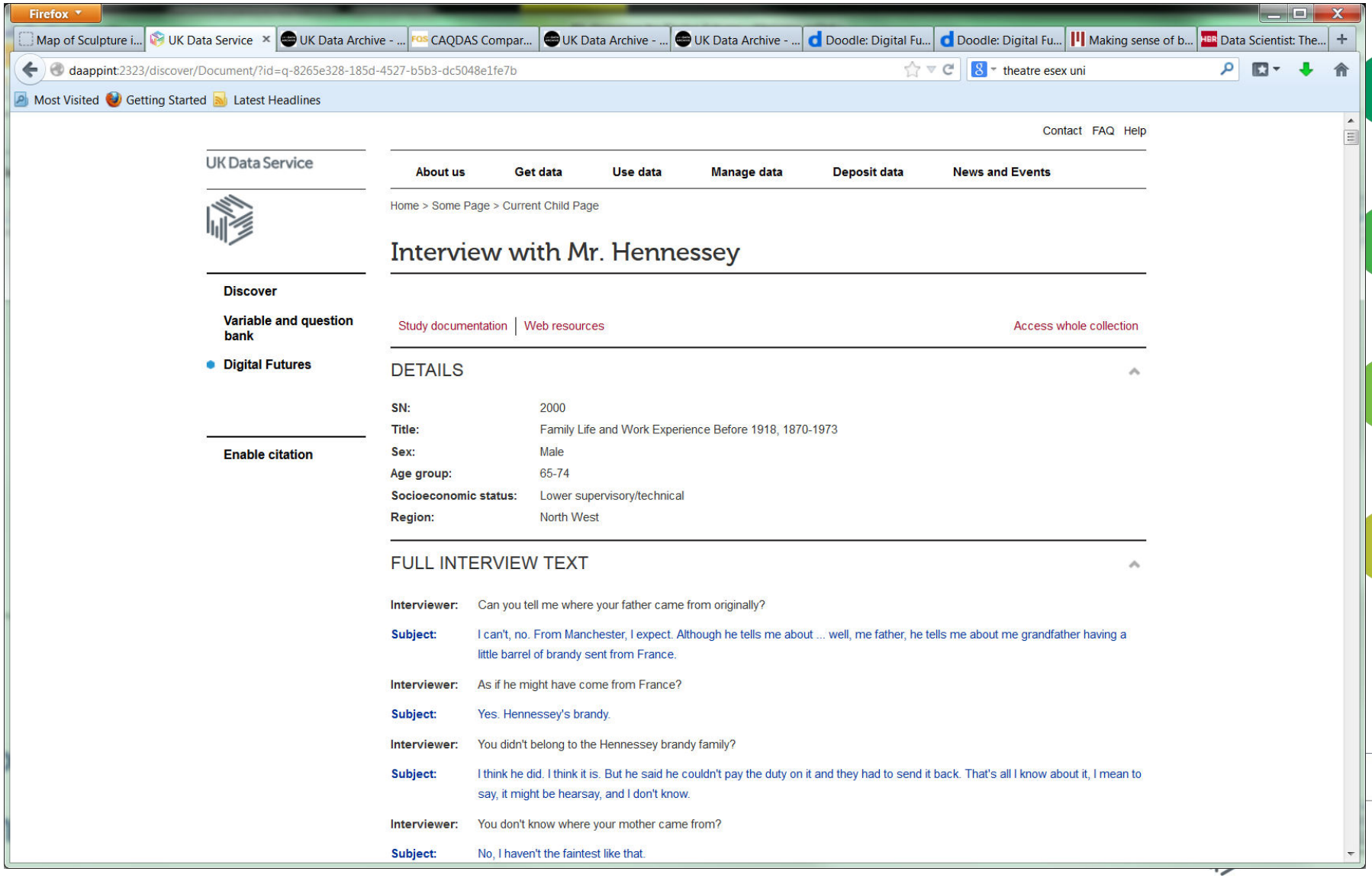
- BaseX for metadata and textual data storage and retrieval
  - DDI 2.5/ DDI Codebook: collection level
  - QuDEx: Limited collection and object level
  - TEI: object level - text documents
- File server for non-XML docs
- Simple QuDEx metadata data entry tool: Sharepoint
- C# scripts process and validate against XML schemas;
- Oxygen for manual mark-up
- Solr indexes used for faceted browsing and TEI text highlighting
- Xquery on BaseX for object metadata, text utterances and related materials
- GUID generator SQL database for QuDEx and TEI elements
- UI in-house technologies .NET and RESTful web services/APIs


Digital Futures Data Flow Diagram
Last Updated – 04 September 2013 - DBell
=PRIMARY MAINTAINABLE REPOSITORY
.NET Console App
Legacy TEI XML
\\daproc4\process$\Digital Futures
\DF collections xml
TEIExtractor
MS-SQL DF Database
TEIUtterances
(Utterance GUID assigned here)
TEICreator
MS-SQL Storm database
Ddielements/maintitle tables
KittyKruncher Console App & XSLT
Publishes DDI2.5 xml
Solr Indexing
Solr Search Database (DF Core)
Xquery to flatten structures for indexing
BASEX
dahdsst2
Publishes QuDEx And TEI XML Files
Master QuDEx Files on Network Folder
\\daproc4\process$\Digital Futures\DF_processed_files\non_TEI_files
Master PDF/Audio/Image Files on Network Folder
\\daproc4\process$\Digital Futures\DF_processed_files\non_TEI_files
Master TEI XML Files on Network Folder
\\daproc4\process$\Digital Futures\DF_processed_files\TEI_files
Validate TEI & x-check With QuDEx
X-check With QuDEx
Export TEI to RTF
RTF Bundles for Ingest
Ingest and DIP
RTF Collection Bundles
Download Packages
Target Page: Utterances and metadata
RESTful Service calling server-side Xquery
Search and Facet
Validate and Load Package
.NET or Java tbc
QGenerator
Generate/Validate QuDEx & X-check with TEI and PDF/Images
Sharepoint DB
dasptest2
Sharepoint QuDEx Data Entry Tool
http://nextgen/df
Publishes non-text files and generates image thumbnails
PDF/Audio/Image Files on Web Server
Display non-text files
Digital Futures UI
Need streamlined workflow!!

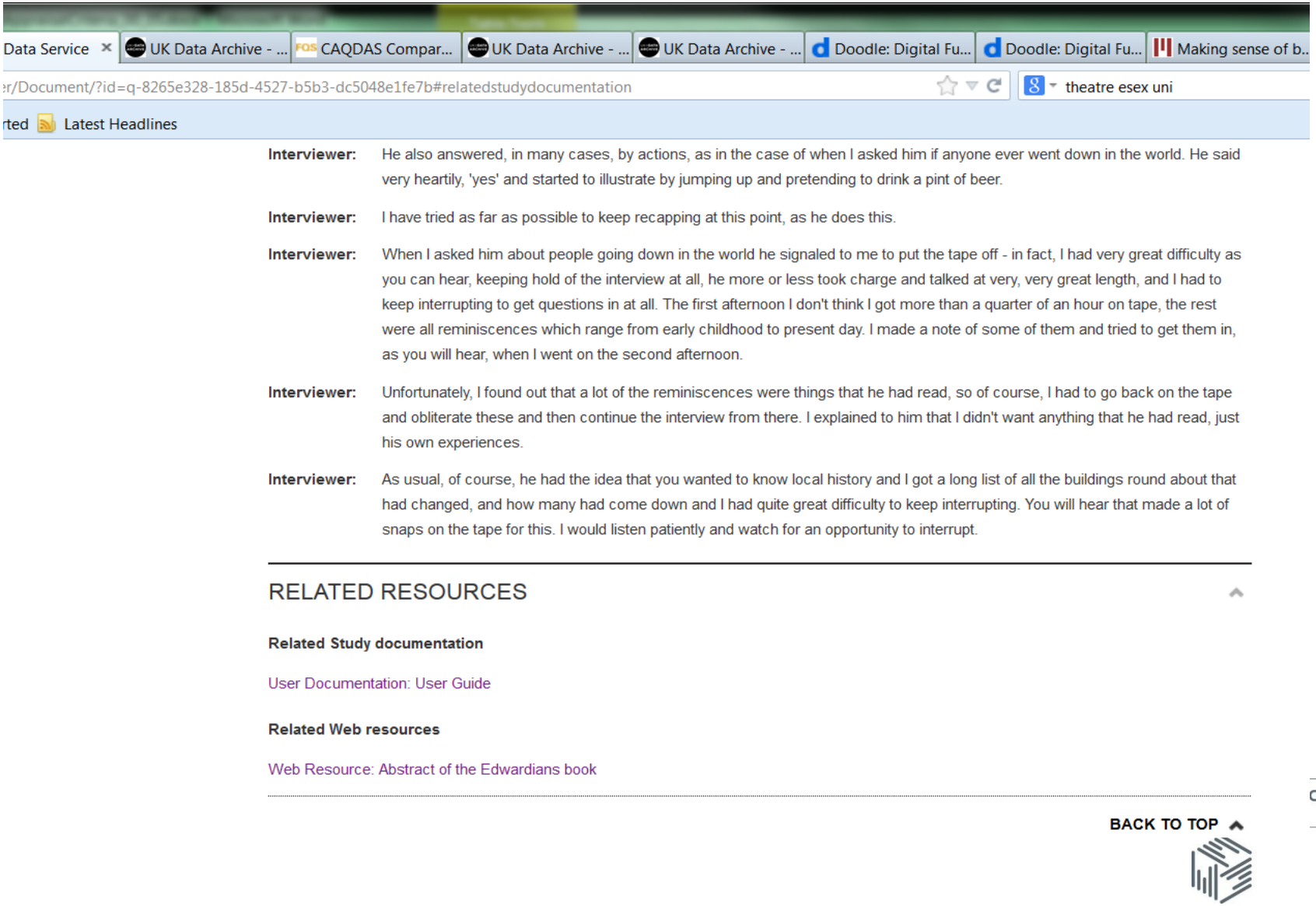
Search interface - hits



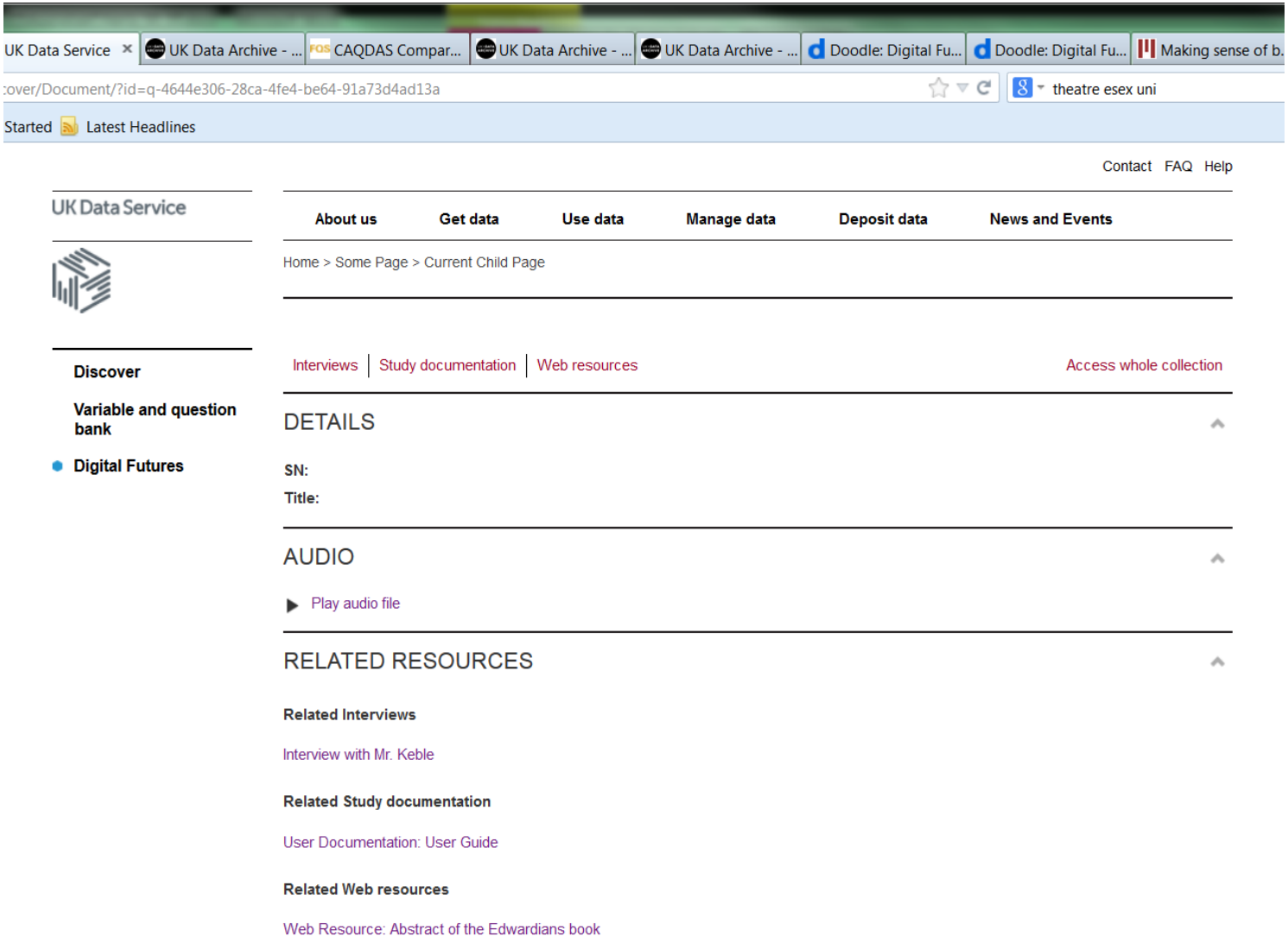
Target page for an interview



Target references



Audio file target page

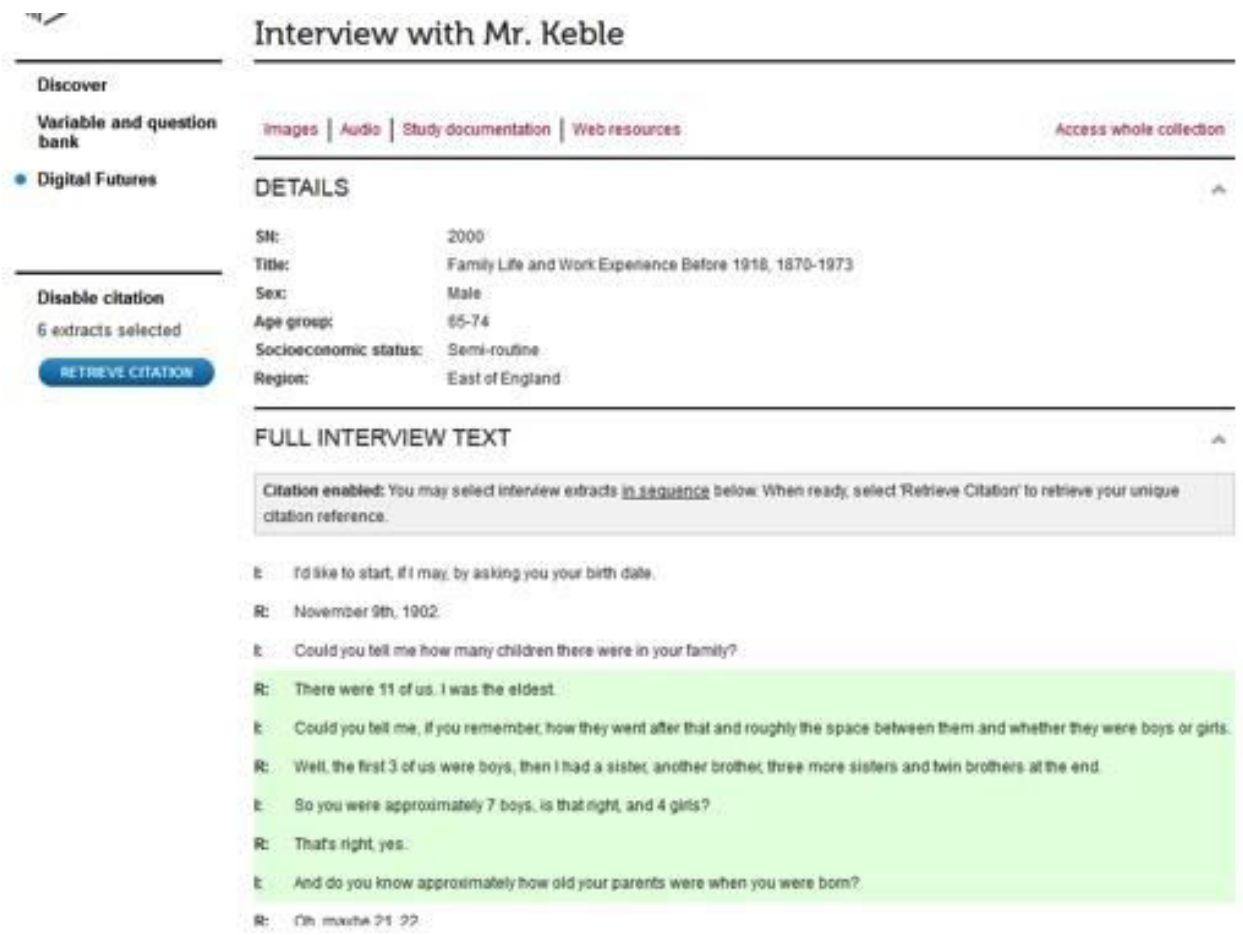


Citation mechanism

- System will enable extract/quotation level citation; 1 or more consecutive paragraphs
- Citation object and citation format created on the fly – using GUIDs and system URI
- URI resolves directly to the data extract
- Some more sensitive collections will be closed, so cannot resolve to data
- Will be related to our collection-level DOIs

10.5255/UKDA-SN-6124-1

Citation



Interview with Mr. Keble

Discover

Variable and question bank

Digital Futures

Disable citation

6 extracts selected

RETRIEVE CITATION

Images | Audio | Study documentation | Web resources

Access whole collection

DETAILS

SN:	2000
Title:	Family Life and Work Experience Before 1918, 1870-1973
Sex:	Male
Age group:	65-74
Socioeconomic status:	Semi-routine
Region:	East of England

FULL INTERVIEW TEXT

Citation enabled: You may select interview extracts in sequence below. When ready, select 'Retrieve Citation' to retrieve your unique citation reference.

I: I'd like to start, if I may, by asking you your birth date.

R: November 9th, 1902.

I: Could you tell me how many children there were in your family?

R: There were 11 of us. I was the eldest.

I: Could you tell me, if you remember, how they went after that and roughly the space between them and whether they were boys or girls.

R: Well, the first 3 of us were boys, then I had a sister, another brother, three more sisters and twin brothers at the end.

I: So you were approximately 7 boys, is that right, and 4 girls?

R: That's right, yes.

I: And do you know approximately how old your parents were when you were born?

R: Oh, maybe 21, 22.

UK Data Service

Questions

Contact details

Louise Corti, corti@essex.ac.uk
Agustina Martinez, a.martinez-garcia@ljmu.ac.uk