
Qudex for Digital Futures: digitizing and publishing qualitative data

Louise Corti

Collections Development and
Producer Support

Digital Futures seminar

FORS, Lausanne, October 2013

UK Data Service



Digital futures project: rationale & aims

- Provide enhanced access to key ESRC-funded qualitative data via online **data browsing and exploration**
- 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



Our initial anticipated audiences

- Researchers: data re-use in activities that inform both research and policy. Expect some Innovative data-rich outputs
- Trainees: capacity building in social science data confrontation and analysis, enhancing data skills
- Citizen researchers: information seeking and knowledge exchange



Users, data needs & enhanced outputs

USER TYPE	DISCIPLINE	DATA TYPES	CONTEXT	ENRICHED OUTPUTS
Academic Researcher	Hot topics	All	Rich	Enhanced Publication
Apprentice Researcher	Hot topics	Standard forms	Rich	Enhanced Dissertation
Teacher/Student	Tried & tested topics	Standard forms	Limited, specific	Enhanced teaching materials/Assessment
Citizen Researcher	Local issues/current affairs/trends	Sample data, simple data – real life issues/'facts'	Little	Own private research/Public visualisation

vice



UK Data Service and its own needs

- We have one of the largest qualitative data collections– over 300 data collections
- A proportion of these have been digitised from older paper sources
- 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
 - It can be a bit of guess work!
 - Cannot reliably cite parts of data



Finding & accessing qualitative data

- Search for “health”
- Retrieve catalogue record, e.g.
 - SN 4943: Mothers And Daughters: Accounts Of Health In The Grandmother Generation, 1945-1978
- View limited user guide
- Web download as RTF bundle (46 interview transcripts)



Clues to content: study level metadata

Morbidity and mortality - Population, vital statistics and censuses
Child development and child rearing - Social stratification and groupings
Elderly - Social stratification and groupings
Family life and marriage - Social stratification and groupings
Gender roles - Social stratification and groupings
Use and provision of specific social services - Social welfare policy and systems

Depositor(s):

Blaxter, M., University of East Anglia

Principal Investigator(s):

Blaxter, M., University of East Anglia

Data Collector(s):

Blaxter, M., University of East Anglia

Sponsor(s):

Economic and Social Research Council

Abstract:

This is an enhanced qualitative study.

The research looked at beliefs and attitudes to **health** and medical care, inter-generational relationships, and social history of members of a grandmother generation. The original study included interviews with daughters as well; this collection contains only the grandmother interviews.

Grandmothers are asked extensive questions about their own **health** and the **health** of other family members. Details are provided on episodes of illness and remedies used, both home and **health** services. Specific topics of accidents, nutrition, dental care, and immunisation are covered.

More generally, grandmothers are asked about their views of their personal doctors and institutional **health** services. They give opinions on the quality of **health** care before and after the introduction of the National **Health** Service.

Grandmother-daughter relationships are explored, especially around the subject of offering and taking of medical advice concerning care for the grandchildren.

The collection has been enhanced by: conversion from paper to searchable RTF format by OCR and extensive editing and formatting of all interview transcripts. This collection will also be made available through [ESDS Qualidata Online](#).

Main Topics:

The interviews cover **health** and social history, beliefs and attitudes to medical care, and intergenerational relationships.

There are discussions of the grandmothers' backgrounds, including jobs they held and where they lived.

Many aspects of **health** are addressed: their own **health**, childhood diseases, the women's attitudes toward doctors and the National **Health** Service, including how medical care changed after the start of the NHS. The grandmothers are asked about causes of diseases, remedies and treatments used, and their thoughts on staying **healthy**.

Relationships with daughters are covered, focusing on whether the grandmothers offer medical advice, and if they do, whether it is accepted by their daughters.

Coverage:

Time Period Covered: 1945-1978

Dates of Fieldwork: 1977-1978

Index date: 1945-1978

Country: Scotland

Spatial Units: No Spatial Unit

Observation Units: Individuals

Kind of Data: Textual data; Semi-structured interview transcripts

Universe Sampled:

Location of Units of Observation: Subnational

Population: Women living in a city in Scotland who had a child between 1950-53, who were at that time in social class IV or V, who had a daughter who: had a child or children born in the same city; and were in the same social classes at the time of the birth; and still lived in the city; and were in touch with the grandmother.

Methodology:

Time Dimensions: Cross-sectional (one-time) study

Sampling Procedures: Purposive selection/case studies

Number of Units: 46

Method of Data Collection: Face-to-face interview; Transcription of existing materials; Audio recording

Weighting: Not applicable

Control Operations:



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



Download Zip

File Name | File Type | Modified | Size | Ratio | Packed | CRC | Attributes | Folder

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...

File Name | File Type | Modified | Size

File Name	File Type	Modified	Size
4943int01.rtf	Rich Text Format	08/07/2004 12:...	108,561
4943int02.rtf	Rich Text Format	08/07/2004 13:...	50,442
4943int03.rtf	Rich Text Format	08/07/2004 13:...	140,718
4943int04.rtf	Rich Text Format	08/07/2004 13:...	168,204
4943int05.rtf	Rich Text Format	08/07/2004 13:...	94,236
4943int06.rtf	Rich Text Format	08/07/2004 13:...	129,875
4943int07.rtf	Rich Text Format	08/07/2004 13:...	73,219
4943int08.rtf	Rich Text Format	08/07/2004 13:...	133,961
4943int09.rtf	Rich Text Format	08/07/2004 13:...	242,754
4943int10.rtf	Rich Text Format	08/07/2004 13:...	120,920
4943int11.rtf	Rich Text Format	08/07/2004 13:...	191,940
4943int12.rtf	Rich Text Format	08/07/2004 13:...	165,688
4943int13.rtf	Rich Text Format	08/07/2004 13:...	197,312
4943int14.rtf	Rich Text Format	08/07/2004 13:...	164,537
4943int15.rtf	Rich Text Format	08/07/2004 13:...	150,199
4943int16.rtf	Rich Text Format	08/07/2004 13:...	161,918
4943int17.rtf	Rich Text Format	08/07/2004 13:...	250,943
4943int18.rtf	Rich Text Format	08/07/2004 13:...	213,834
4943int19.rtf	Rich Text Format	08/07/2004 13:...	159,213
4943int20.rtf	Rich Text Format	08/07/2004 13:...	135,742
4943int21.rtf	Rich Text Format	08/07/2004 13:...	196,686

Service



Typical rft transcript template

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 after my grandchildren. Like that... well, my grand-daughter comes in at
night, well, her mither cleans the school... and I look after 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 after 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.

Complex data collections

- SN 5801: Concepts of Healthy Eating Food Research: Phases I and II, 1992-1996
- 293 interview transcripts; 73 diaries; 6 observation 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



Satisfying modern-day users

- 21st Century system for delivering qualitative data
- Web-based data exploration system
- Interface should lower the barriers to non-academic users
- Explore data through a data journey
 - Find relevant **extract, examine in context, cite**
 - **Link data** to still and moving images, and other related research outputs



Metadata demands for the system

- Demands highly structured and consistently marked-up data
- Qualitative data requires object (file-level) descriptive metadata, e.g. interviews, audio-visual files, images
- Common metadata elements enable federated catalogues across providers and borders



Richer metadata = richer discovery

- Enabling rich description
- Use of DDI 2.5, QuDEX and TEI schema
- QuDEX allows identification of data objects:
 - Interview transcript or audio recording etc.
 - Relationship to another data object or part of data
 - Descriptive categories at the object level, e.g. mime type, interview characteristics, interview setting
 - Capacity to capture rich annotation of parts of data
- Qudex model in use (Schema at: www.data-archive.ac.uk/create-manage/projects/qudex/)
- Object-level description = a lot of manual work!



System expectations

- **Search/browse for data**
 - Search /faceted browse of data - text; image/PDF, audio
- **Browse**
 - Faceted browse by categories:
 - Collection level, title, date and openness
 - Collection object: 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
 - Get citation for part of data



Search interface - hits

The screenshot shows a Firefox browser window displaying the 'Digital Futures' search interface. The browser's address bar shows the URL 'daappint:2323/discover/DigitalFutures'. The page features a search bar with the text 'illness' and a 'GO' button. Below the search bar, there are filters for 'Utterance', 'Image', 'File', and 'Audio'. The search results are displayed under the heading 'SEARCH RESULTS' and show 'Displaying 1-10 of 235 results'. The first result is 'Interview with G18' from 'SN4943 Mothers and Daughters, 1945-1978', with a snippet: '...But you had a lot of illness when you were at school... uh-huh...'. The second result is 'Interview with G32' from the same collection, with a snippet: '...Oh, I'd lots of illnesses then. You know, I've had diphtheria, scarlet fever, every illness you could think of. Pneumonia, double pneumonia, everything. And then I had a ... the palette of my throat fell down wif diphtheria, an' I was dumb for six months. I still remember that. I wis just about...'. The third result is 'Interview with G7' from the same collection, with a snippet: '...'. Each result includes a 'Refine' button and a link to 'Access whole collection'. The left sidebar contains navigation options: 'Discover', 'Variable and question bank', 'Digital Futures', 'HASSET', and 'ELSST'. Below these are filters for 'Collection title' and 'Resource type'.

Firefox

daappint:2323/discover/DigitalFutures

Digital Futures

Search and browse qualitative surveys, interviews and open-ended questions

illness GO

Reset filters | Clear search | Auto-complete | Help

Utterance | Image | File | Audio

SEARCH RESULTS

Displaying 1-10 of 235 results

1 2 3 4 5 >>>

Interview with G18
SN4943 Mothers and Daughters, 1945-1978
Sex: Female. **Age group:** 55-64. **Socioeconomic status:** Never worked/long-term unemployed. **Region:** Scotland.

...But you had a lot of **illness** when you were at school... uh-huh...

[Access whole collection](#)

Interview with G32
SN4943 Mothers and Daughters, 1945-1978
Sex: Female. **Age group:** 55-64. **Socioeconomic status:** Never worked/long-term unemployed. **Region:** Scotland.

...Oh, I'd lots of illnesses then. You know, I've had diphtheria, scarlet fever, every **illness** you could think of. Pneumonia, double pneumonia, everything. And then I had a ... the palette of my throat fell down wif diphtheria, an' I was dumb for six months. I still remember that. I wis just about...

[Access whole collection](#)

Interview with G7
SN4943 Mothers and Daughters, 1945-1978
Sex: Female. **Age group:** 55-64. **Socioeconomic status:** Never worked/long-term unemployed. **Region:** Scotland.

Discover

Variable and question bank

• Digital Futures

HASSET

ELSST

Collection title

The Edwardians (149)

100 Families (64)

Mothers and Daughters (22)

[Refine](#)

Resource type

Interview transcript (234)

Interview summary (1)

Annotated questionnaire (0)

User Documentation (0)

Web resource (0)

audio (0)

image (0)

[Refine](#)

Target page for an interview

The screenshot shows a Firefox browser window with the URL `daappint:2323/discover/Document?id=q-8265e328-185d-4527-b5b3-dc5048e1fe7b`. The page is the UK Data Service website, featuring a navigation menu with links for 'About us', 'Get data', 'Use data', 'Manage data', 'Deposit data', and 'News and Events'. The main content area is titled 'Interview with Mr. Hennessey' and includes a 'DETAILS' section with the following information:

SN:	2000
Title:	Family Life and Work Experience Before 1918, 1870-1973
Sex:	Male
Age group:	65-74
Socioeconomic status:	Lower supervisory/technical
Region:	North West

Below the details is a 'FULL INTERVIEW TEXT' section containing the following dialogue:

Interviewer: Can you tell me where your father came from originally?
Subject: I can't, no. From Manchester, I expect. Although he tells me about ... well, me father, he tells me about me grandfather having a little barrel of brandy sent from France.

Interviewer: As if he might have come from France?
Subject: Yes. Hennessey's brandy.

Interviewer: You didn't belong to the Hennessey brandy family?
Subject: I think he did. I think it is. But he said he couldn't pay the duty on it and they had to send it back. That's all I know about it, I mean to say, it might be hearsay, and I don't know.

Interviewer: You don't know where your mother came from?
Subject: No, I haven't the faintest like that.

Target references

The screenshot shows a Firefox browser window with the following elements:

- Browser Tabs:** Map of Sculpture i..., UK Data Service, UK Data Archive - ..., CAQDAS Compar..., UK Data Archive - ..., UK Data Archive - ..., Doodle: Digital Fu..., Doodle: Digital Fu..., Making sense of b..., Data Scientist: The...
- Address Bar:** daappint:2323/discover/Document/?id=q-8265e328-185d-4527-b5b3-dc5048e1fe7b#relatedstudydocumentation
- Page Content:**
 - Interviewer:** He also answered, in many cases, by actions, as in the case of when I asked him if anyone ever went down in the world. He said very heartily, 'yes' and started to illustrate by jumping up and pretending to drink a pint of beer.
 - Interviewer:** I have tried as far as possible to keep recapping at this point, as he does this.
 - Interviewer:** When I asked him about people going down in the world he signaled to me to put the tape off - in fact, I had very great difficulty as you can hear, keeping hold of the interview at all, he more or less took charge and talked at very, very great length, and I had to keep interrupting to get questions in at all. The first afternoon I don't think I got more than a quarter of an hour on tape, the rest were all reminiscences which range from early childhood to present day. I made a note of some of them and tried to get them in, as you will hear, when I went on the second afternoon.
 - Interviewer:** Unfortunately, I found out that a lot of the reminiscences were things that he had read, so of course, I had to go back on the tape and obliterate these and then continue the interview from there. I explained to him that I didn't want anything that he had read, just his own experiences.
 - Interviewer:** As usual, of course, he had the idea that you wanted to know local history and I got a long list of all the buildings round about that had changed, and how many had come down and I had quite great difficulty to keep interrupting. You will hear that made a lot of snaps on the tape for this. I would listen patiently and watch for an opportunity to interrupt.
- RELATED RESOURCES**
 - Related Study documentation**
 - User Documentation: [User Guide](#)
 - Related Web resources**
 - Web Resource: [Abstract of the Edwardians book](#)
- BACK TO TOP** (with upward arrow icon)
- FUNDED BY**
 - ESRC ECONOMIC AND SOCIAL RESEARCH COUNCIL
- SUPPORTING**
 - CESSDA Council of European Social Science Data Archives
- NEW USERS**
 - Comni diaecti atlati atios
 - Nihicil id ulluptate consed mo tem
- MOST POPULAR PAGES**
 - Nostimo averehebus
 - Seripio potis egere ces ad tiam
- CONNECT WITH US**
 - Contact (envelope icon)
 - Twitter (bird icon)
 - Facebook (f icon)

Audio file target page

Firefox

Map of Sculpture i... UK Data Service x UK Data Archive - ... FOS CAQDAS Compar... UK Data Archive - ... UK Data Archive - ... Doodle: Digital Fu... Doodle: Digital Fu... Making sense of b... Data Scientist: The...

daappint:2323/discover/Document/?id=q-4644e306-28ca-4fe4-be64-91a73d4ad13a

theatre esex uni

Most Visited Getting Started Latest Headlines

Contact FAQ Help

UK Data Service

About us **Get data** **Use data** **Manage data** **Deposit data** **News and Events**

Home > Some Page > Current Child Page

Discover

Variable and question bank

● Digital Futures

[Interviews](#) | [Study documentation](#) | [Web resources](#) [Access whole collection](#)

DETAILS

SN:

Title:

AUDIO

▶ [Play audio file](#)

RELATED RESOURCES

Related Interviews

[Interview with Mr. Keble](#)

Related Study documentation

[User Documentation: User Guide](#)

Related Web resources

[Web Resource: Abstract of the Edwardians book](#)

FUNDED BY

NEW USERS

Comni diaecti atiatl atios

MOST POPULAR PAGES

Nostimo averehebus

CONNECT WITH US


[Email](#) [Twitter](#) [Facebook](#)

System assumptions

- BaseX for metadata storage; Java loading; Solr search
- Data must be fully prepared on loading/publishing to the system. Data not 'managed' within the system
 - Mark-up, metadata, relationships all pre-defined
 - Pre-defined GUIDs to be used for citation (DOI + drilldown)
- Cannot search audio-visual data content
- Simple QuDEX metadata data entry tool created using Sharepoint
- Technologies for user interface are .NET
- No download of data collection/subset - route to the UK Data Service
- Citation of selected extract of text; user-annotation possible



QuDEX collection level metadata



UK DATA ARCHIVE Digital Futures > QuDEX Headers > All Items

Search this site...

ID	Display Label	Collection Description	Study Number	Access	Status	Modified	Modified By
1	The Affluent Worker	Affluent Worker in the Class Structure: a Digitised Sample of the Luton Study, 1961-1962	4871	open	Draft	28/08/2013 11:06	Bell, Darren S
2	The Edwardians	Family Life and Work Experience Before 1918, 1870-1973	2000	open	Draft	28/08/2013 11:07	Bell, Darren S
3	Poverty in the UK	Poverty in the United Kingdom : a Survey of Household Resources and Standards of Living, 1967-1969	1671	open	Draft	28/08/2013 11:07	Bell, Darren S
4	School Leavers Study	School Leavers Study, 1978	4867	open	Draft	28/08/2013 11:08	Bell, Darren S
14	Mothers Alone	Mothers Alone : Poverty and the Fatherless Family, 1955-1966	5072	open	Draft	28/08/2013 11:08	Bell, Darren S
15	100 Families	Families, Social Mobility and Ageing, an Intergenerational Approach, 1900-1988	4938	open	Draft	28/08/2013 11:08	Bell, Darren S
16	Mothers and Daughters	Mothers and Daughters : Accounts of Health in the Grandmother Generation, 1945-1978	4943	open	Draft	28/08/2013 11:09	Bell, Darren S
17	Angels in Marble	Angels in Marble	1204	open	Draft	27/09/2013 13:55	Bell, Darren S
18	Adultery	Adultery - by Annette Lawson	4858	open	Draft	27/09/2013 10:59	Bell, Darren S

[Add new item](#)



Objects in collection metadata

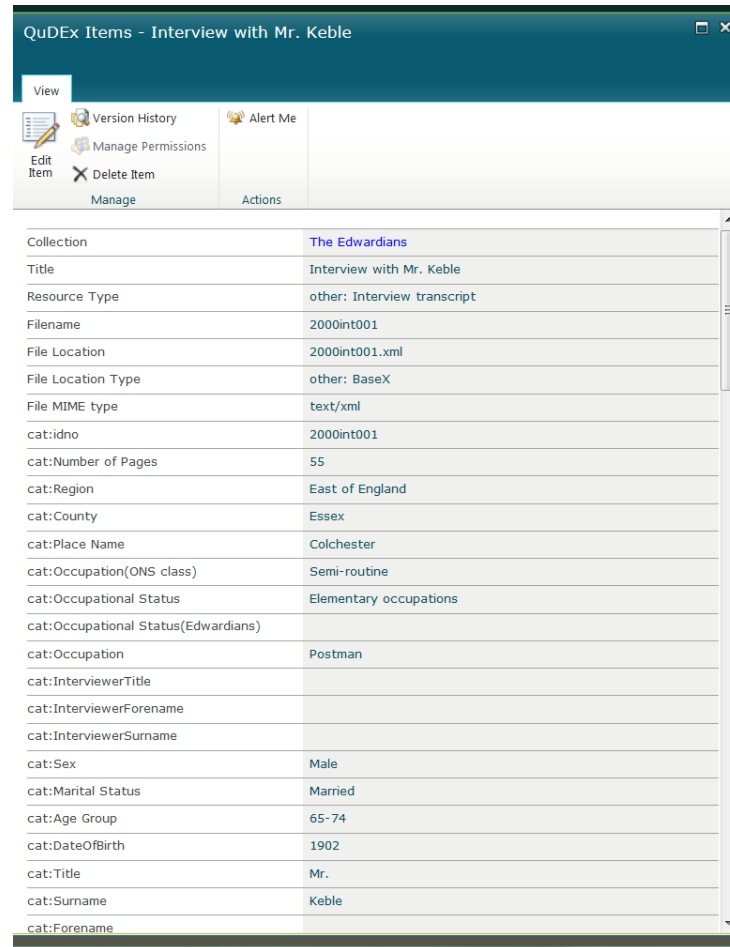
UK DATA ARCHIVE Digital Futures > QuDEX Items > Data Entry Grid =

Search this site...

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



The screenshot shows a web application window titled "QuDEx Items - Interview with Mr. Keble". The interface includes a menu bar with options like "View", "Version History", "Alert Me", "Edit Item", "Manage Permissions", "Delete Item", and "Manage". Below the menu is a table of metadata for the item "Interview with Mr. Keble".

Property	Value
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	



Digitisation of key data sources

Selectively digitize paper-based materials:

- original survey questionnaires
 - interviews
 - handwritten field notes, essays
 - diagrams
 - photographs
-
- Wide consultation to choose sources: user-driven, not archival
 - Large digitization budget



Digitising textual data

Create image file

- used for poor typeface, handwritten, text with tables and graphs
- anonymise a copy with black marker prior to scan
- [scan and save as TIFF image file](#)

Create searchable PDF

- collate TIFFs and convert to PDF (Adobe Acrobat)
- bookmark PDF file for navigation, with contents page and headings

Create rich text: optical character recognition (OCR)

- convert TIFF to RTF format
- requires detailed checking and proofreading
- [XML mark-up using simple TEI profile](#)

Transcription

- Selective transcription method
- create consistent structure; templates
- XML mark-up using TEI

Timelines and deliverables: 18 months

Nov 12 – March 13

- Requirements gathering and technical specification work, making use of project consultants and users
- Mark-up of test data collections in Qudex
- Mark-up of test XML example text in TEI

April – June 13

- Metadata and data (XML) mark-up defined and signed off
- Metadata mark-up tool developed
- Tech spec for system finalised
- Collections scoped and selected for digitisation

Timelines and deliverables

June – January 13

- Data digitization and mark-up of 10-15 collections
- Data loading system and user interface build
- Investigate methodology for permanent citation of data extracts

December 13 – Feb 14

- Develop citation mechanism
- Work with key users to create outputs; teaching and research
- Create an enhanced publication linking to the live system
- Launch system with data – open and closed



Questions

Louise Corti

corti@essex.ac.uk

