

Enhancing Access to Qualitative Data Resources: Edwardians Online

Emma J. Barker; Kevin Schurer; Louise Corti
Qualidata, UK Data Archive, University of Essex



6/9/02

<http://edwardians.qualidata.ac.uk/>

Edwardians Online Project

- Freely accessible web resource based upon 500 life history interviews, Paul Thompson's *Family Life and Work Experience Before 1918*.
- Undertaken by Qualidata in conjunction with UK Data Archive
- On-going initiative to enhance access to qualitative data resources
- Encouraging reuse of qualitative data

Outline of Talk

- 6 month pilot project near to completion
- Background: original study and resources
- Outline of problem underlying re-use of data
- Description of online project:
 - Aims and objectives
 - On-line deliverables
 - Methodology: tools and data formats
 - Problems encountered
- Further applications and directions for future research

"Family Life and Work Experience Before 1918"

- Dataset of life history interviews; classic sociological study of Edwardian Society by Professor Paul Thompson
- Conducted in early seventies over 3 years; nearly 500 interviews completed
- Cross-national sample of people born in Britain before 1918
- Originally recorded on audio tapes; transcribed as typed documents; texts coded in thematic analysis of content
- Dataset archived, catalogued and disseminated by Qualidata

- Interviews up to 100,000 words
- 8hrs of audio tape

Loosely structured dialogue –

- alternate speakers
- pre-specified interview schedule

- 42 -

of that. But it was - you called them motor caps. I don't ken what it was for and that but I can always mind of that.

But you weren't actually at your wee sister's funeral then?

Oh no, no, I woldna - you ken, I dinna - couldna say that I can mind her even. I cann mind her even being, ken, took out of the house.

Did you rememer anybody else's funeral then? Maybe your grandparents or even somebody in the street that maybe everybody went to?

Well tell you the truth and that, even - you ken, funerals even at the present day, they dinna, ken, appeal to me. It's - it's just I would rather not - even my sister in Brooms Road, well I wasnae able - to, ken, to have gone, but even - ken, I sit and I many a time think here to - I would rather - and even Johnny up the road, like my brother that died last year, the year before, I mean, even - I wasna there either. But I think I would rather just - think on them as I last seen them ken like. I - I - I mean - I just dinna ken, I just dinna -

Interview Schedule

1. The Household

a) Respondent's name, present address, year of birth, marital status, year of marriage, birthplace (street or district if known).

b) How many years did you live in the house where you were born? Where did you live then? CONTINUE FOR MOVES TO END OF 1918. FIND OUT ADDRESS AS NEARLY AS POSSIBLE FOR 1911.

Do you remember why the family made these moves?

c) How many brothers and sisters did you have? Birth order and spacing.

d) How old was your father when you were born? (PROMPT: How old was he when he died? When was that?)

Where did he come from?

Occupation. (IF EMPLOYER: How many people did he employ?)

Did he have another job before or after he became that?

Did he also do any casual or part-time jobs?

CONTINUE FOR ALL JOBS UNTIL DEATH, INCLUDING AFTER 1918.

Do you remember your father ever being out of work?

Interview
schedule

Thematic codes

- Questions loosely grouped in these themes
- Themes used to classify segments of text
- May be overlapping

1. Household
2. Domestic routine
3. Meals
4. Influence and discipline
5. Recreation in the home
6. Recreation outside the home
7. Weekend activities and religion
8. Politics
9. Parents' interests
10. Children's leisure
11. Community and social class
12. School
13. Work, except domestic service
14. Life after leaving school
15. Marriage
16. Childbirth - including sexual knowledge
18. Domestic service
19. Institutions and boarding schools

Reuse of Qualitative Data: Access Constraints

- Large scale qualitative studies -such as Edwardian interviews- have great reuse potential
- Manual access to paper resource and finding aid time-consuming; expensive
- Searching by new topic or key word impractical
- Data dispersed in various formats and different locations
- Ideally want content based access

Online Project-aims and objectives

- Digitally formatted, web-based, multi media resource based on original collections for research and teaching
- Integrating primary and secondary materials
- Providing content based search and retrieval tools
- Integrate metadata with content model
- To allow for interoperability and data-interchange by providing resource in portable and long term, non proprietary format

Online Project-Deliverables

- Electronic catalogue of interview summary data
- Indexed database of electronic transcripts
- Search and browse tool for thematic extracts of text – linked to full document context
- Database of short sound clips
- Background material and resources relating to original project and context
- Information on reuse and secondary analysis

Methodology: Tools and Data Formats

- Bottleneck in input processing –need for speed and control in creation of digitally formatted transcripts from paper texts
- Data Entry: scanning and OCR; additional proofing
- Semi-automated pre-processing for adding structure to text
- Pilot project adopted database methodology
- RDBMS: supports data management; straightforward data input; coding; on-line searching/retrieval
- Relational model appropriate for integrating metadata, document content and sound clips; thematic annotation

line_id	doc_id	u_id	who	utterance
171	200	173	inter	Did anyone else besides your parents and brothers and sisters live in the house, did you have any relatives or lodgers living with you?
172	200	174	subj	We had my grandma came.
173	200	175	inter	How old would you be when your grandmother came to live with you?
174	200	176	subj	Well I should only - think I would only be about five year old.
175	200	177	inter	How long did your grandmother live with you?
176	200	178	subj	Oh dear - she - until our Margaret was born and she was - the last one.
177	200	179	inter	So she lived with you quite a long time?
178	200	180	subj	Yes, she did. Oh Margaret's - I can tell you Margaret's age. She's sixty-one.
179	200	181	inter	So that was 1905 - 1909?

line_id	theme_id
171	2
171	11
172	2
172	11
173	2
174	1
174	2
175	1
175	2
176	1
176	2
177	1
177	2
178	1
178	2
179	1
179	2

Coding of utterance (turn of speech) by theme

XML and Standard Archive Formats

- XML applications emerging as a standard archive format for metadata and also the data itself
- Pilot study initiated research into comprehensive DTD for qualitative datasets (metadata and content) using existing standards
- TEI guidelines for transcriptions of speech provide basis for content model (interview structure)
- DDI framework provides basis for header and metadata
- Aim: To map interview database into XML for preservation format

Problems

- Multiple hierarchies and overlapping fields:
Structure v thematic content
- Violates nesting rules of XML
- Relational model accommodates this easily
- Similar model in XML : “stand-off annotation”

Conclusions and Future Research

- Full functionality pilot resource to be completed June 2002
- Develop full DTD for Qualitative Data-comprehensive metadata;document structure;thematic analysis
- Importance of stand-off annotation model
- Incorporate within DDI