

Edwardians Online

Introduction

In June 2002, the Qualitative Data Service (Qualidata) based at the UK Data Archive released Edwardians Online, a pilot, web-based, multimedia resource. The aim of this work is to develop a standard framework for digitizing data collections, and to provide online access to the **content** of digitized qualitative data collections. Edwardians Online integrates a wealth of existing primary and secondary materials relating to an oral history study carried out in the 1970s. Researchers can perform free-text and thematic database searches of the interview summaries, as well as a sample of

By *Emma J Barker*
&
*Louise Corti**

the full text transcripts. Thematic searches are based on the existing coding schema originally used to classify and analyse the data. Linked to this primary material are sound extracts from the audio recordings, images and contemporary photographs. Further background material relating to the original research study, such as press reviews and details of publications based on secondary studies of the interview texts are also included. Phase II of the project aims to link other key sources, including maps and census data.

The screenshot shows the homepage of the Edwardians Online website. At the top left is the logo "Edwardians Online" in a blue script font, with the tagline "FAMILY LIFE AND WORK EXPERIENCE BEFORE 1918" underneath. To the right of the logo is a black and white photograph of a group of people. Below the logo and photo is a navigation bar with links: "About | Search | Original Project | Re-Use | Images | Sound". On the left side, there is a vertical menu under the heading "About" with links for "Introduction", "Online Project", "Citation Information", "Contact Details", and "Feedback and Evaluation". The main content area is titled "Introduction to Edwardians Online" and contains several paragraphs of text. The first paragraph states that the project aims to improve access to archived, qualitative data. The second paragraph describes the interviews as part of Professor Paul Thompson's study of Edwardian society. The third paragraph mentions that Phase I of the project is complete and lists several features available online, such as a catalogue of 444 interview summaries, full electronic transcripts, thematic browsing, digital sound clips, contextual images, background information, and details of publications. The final paragraph notes that the original study materials were archived by Qualidata and are held at the National Social Policy and Social Change Archive at the University of Essex. At the bottom of the page, there is another navigation bar with links: "Top of Page | Home | Copyright and Disclaimer | Help? | Site Map" and "You are here: Edwardians Online > Introduction to Edwardians Online". Below this are logos for Qualidata, UK Data Archive, University of Essex, and the University of Essex logo. At the very bottom, there is contact information for Qualidata at the University of Essex, including the address, telephone number, fax number, and email address, along with a copyright notice for 2002.

Increasing Access to Qualitative Data: Digital Resources

Qualidata provides a national service for the acquisition, dissemination and re-use of qualitative data that have been collected using social science research methods. The issue of how to make these data resources accessible to users is a central concern for the service. We are continually seeking ways to meet users' requirements. Results of earlier work in this area can be seen in Qualidata's resource discovery hub, where users can search and locate accessible collections of qualitative data across the UK via the online catalogue, Qualicat. More recently, in response to user demand, emphasis has been on data development, with a view to providing users with **direct access** to the **content** of digitized collections via an online facility.

The first steps in this content-oriented direction include an increased focus on depositing digital data in-house with the UKDA, and the digitization of 'classic collections' for research and teaching resources. Edwardians Online, based upon a set of oral history interviews, was selected as an appropriate collection for undertaking Qualidata's first major web-based digitization project.

The Data Collection

The interviews were undertaken in the early 1970s as part of Professor Paul Thompson's study of Edwardian society. These interviews form the basis for Thompson's, *The Edwardians: The Remaking of British Society*, (1975, 1992). The 444 interviews were originally recorded on audio tapes and later transcribed as typed, paper documents. The original study materials were initially archived, catalogued and disseminated by Qualidata.

The importance of this collection for secondary use lies in the diversity and broad scope of the interview content and the scale of the collection. The collection has attracted high usage across a variety of research interests and is a valuable teaching resource. Users have requested access to both complete interview transcripts and more specific information or extracts from within the documents. Due to the length of the interviews, use of the collection can be time-consuming. For example, a typical transcribed interview may be 80 typed pages and an audio recording as long as four hours.

Phase I of the Project

A main aim of this project is to produce a prototype methodology which may be developed into a more general application for other examples of social science datasets. Research on this collection has focused on the following key areas:

- developing a non-proprietary electronic format for preserving the content of qualitative datasets.
- developing tools for facilitating the encoding of data in this format

- questioning the methods of access and facilities for exploring qualitative data online

A Standard Framework for Archiving Digital Qualitative Data Resources

In pursuit of these issues, a comprehensive application appropriate for interchange that will enable sophisticated on-line searching and information retrieval from encoded texts is required. Ideally, the application should meet a number of specific objectives that:

- support the encoding of the content of various types of primary data documents produced in qualitative research.
- support the encoding of contextual documentation and metadata linked to the primary sources
- provide formalised links between the texts and associated audio and video materials, with a view to providing in the long term, integrated, multi media resources.
- represent the content of datasets, such as the researcher's original analytic schema, annotations and speaker tags.

A uniform format for encoding the content of datasets is useful for both data providers and users. It ensures consistency across datasets; supports the development of common publishing and search tools; and facilitates data interchange and comparison between datasets.

Development of an XML Application for Qualitative Data

Finding a framework that will enable these functions leads us to consider XML standards and technologies. XML and related tools for creating and processing documents in XML have rapidly been adopted by communities of users for whom semantic tagging for their own application areas is essential. Examples where XML tag sets are specially adapted to allow markup of the types of information specific to the user community include the Data Documentation Initiative (DDI) for the social sciences and the Text Encoding Initiative (TEI).

With increasing recognition of the benefits of XML in creating non-proprietary, cross-platform applications, there has been serious interest in, and calls for, the development of a qualitative data XML markup language from members of the social science research community who are eager to encourage the re-use of social science data.

The development of a common framework for marking up the content of qualitative datasets requires support and contributions from various members of the social science community: data creators; qualitative data software

developers; data providers and end users. In particular agreements need to be made on:

- types of documents and structures to be marked up
- formal definition of a common XML vocabulary and DTD for describing these structures
- specification of publishing and analysis tools
- test applications with 'real' datasets

Edwardians Online has aimed to provide the foundations for a broader initiative. Research to date considered two options: first, to create a customized application of XML specifically for the purpose of marking up the content of spoken interviews and other types of qualitative material. Second, to adapt existing standards, such as the TEI and the DDI, thereby opening opportunities for using existing and forthcoming tools for processing the XML texts, in addition to the benefits of using a standard, such as detailed documentation and the expertise and experience of the previous user community.

Conclusion

These ideas will be explored further in Phase II of the project. Phase II will focus on the development of additional search and retrieval functionality, and the encoding of additional features in the interview texts. The presentation of a Document Type Definition (DTD) for a generalized XML application for qualitative datasets is a key milestone in this programme. Over the next year we will begin work on adapting and integrating the TEI and DDI to produce a prototype DTD for qualitative data. We hope this would become a de facto standard and one that could be used by other data creators and data publishers to encode a broad class of qualitative data.

We are keen to encourage testing of the pilot resource. If you would like to complete a simple user evaluation exercise, please go to the web site and look for the Evaluate tab. The resource can be found at: <http://www.qualidata.essex.ac.uk/Edwardians>. For further details of ongoing and future work contact **Louise Corti** at the UK Data Archive.

Acknowledgement

Emma Barker, the project officer, left the UK Data Archive in December. The UKDA would like to **extend its thanks to Emma for the** dedication she has shown to Phase I of this project.

References

DDI <http://www.icpsr.umich.edu/DDI/ORG/index.html>

Qualidata, <http://www.qualidata.essex.ac.uk/>

National Social Policy and Social Change Archive, <http://www.qualidata.essex.ac.uk/dataResources/nspsca.asp>

National Sound Archive, the British Library,

<http://www.bl.uk/collections/sound-archive/holdings.html#story>

TEI www.tei-c.org/

Thompson, P.R. (1975), *The Edwardians, The Remaking of British Societ*. London: Granada

* Paper presented at the IASSIST Conference, June 2002, in Storrs, CT, USA. Louise Corti is Associate Director & Head, Qualitative Data Service and Outreach and Training at the UK Data Archive, University of Essex, email: cortl@essex.ac.uk