Infrastructure services and needs for the provision of enhanced qualitative data resources*

Q1

Louise Corti

Introduction

The aim of this paper is to provide an overview of the opportunities and challenges faced by a national qualitative data service provider, and specifically, how these relate to broader infrastructural requirements.

In the first part of the paper I provide an overview of past and current qualitative data provision, largely with reference to the UK position. This is followed by a look at new directions in data provision and in related support services. Tracing the history of the UK service Qualidata since 1994 enables us to pinpoint how major cultural and funding barriers have been overcome, and how new opportunities have enabled the UK service to gain a renewed lease of life.

Louise Corti is Associate Director and Head of the Qualitative Data Service and Outreach and Training at the UK Data Archive, University of Essex, UK. In the past she has taught sociology and social research methods, and spent six years working on the design, implementation and analysis of the British Household Panel Study at the University of Essex. She is interested in both qualitative and quantitative aspects of social research.
Email: cortl@essex.ac.uk

In the second part of the paper I outline the new UK Economic and Social Data Service, which has a specific qualitative data component built in to it. I describe the new directions and promised deliverables of the service, and suggest how making these successful will be greatly aided by some key developments in national infrastructure. Two such areas are the establishment of more coordinated and mandatory social science datasets policies by research funders; and a nationally coordinated research methods training strategy that recognises the value of secondary analysis of qualitative data. The issues are further illustrated by drawing on the example of a pilot project to make qualitative research data on Edwardian England available online.

Qualidata: history

In the UK, until recently, no infrastructure existed for the systematic archiving and dissemination of qualitative data from social science research. The Economic and Social Research Council (ESRC) had already recognised, very early on in 1967, the value in retaining the most significant machine-readable data from the empirical research that it funded by establishing a Data Archive. Since the 1970s, social science data archives across the world have typically acquired a significant range of data relating to society, both historical and contemporary from sources including surveys, censuses, registers and aggregate statistics. Equally, these centres of expertise have established networks of data services for the social sciences that foster cooperation on key archival strategies, procedures and technologies.

Thus crucial survey data can be re-analysed by other researchers, and the money spent on research has become not only an immediate outlay but also an investment for the future. There was, however, a significant gap in this policy in that qualitative data were rarely

ISSJ 177 © UNESCO 2003. Published by Blackwell Publishing Ltd., 9600 Garsington Road, Oxford OX4 2DQ, UK and 350 Main Street, Malden, MA 02148, USA.

acquired, in spite of much data being created in word-processed form. The 1990s saw a growing demand for access to digital texts, images, and audio-visual material. When a small pilot study commissioned by the ESRC was carried out by Paul Thompson in 1991 (Thompson 1991), it was revealed that 90% of qualitative research data was either already lost, or at risk, in researchers' homes or offices. Moreover, the 10% "archived" was found not to have the basic requirements of an archive, such as physical security, public access, reasonable catalogues, with recorded material or listening facilities. It was further calculated that it would cost at least £20 million to create a resource on the scale of the data at risk. For the older British sociological material, moreover, the risk was acute, and the need for action especially urgent. This was subsequently borne out by the destruction of research data from the classic UK community studies of Banbury (Stacey 1974) and Sparkbrook (Rex & Moore 1967), and the Q2 UK longitudinal study on child-rearing by John and Elizabeth Newson (1976).

In 1994, the first qualitative data archiving project on a national scale was established in the UK, with support from the ESRC. Housed within the Department of Sociology at the University of Essex, its objectives were to facilitate and document the archiving of qualitative data arising from research, whilst also drawing the attention of research communities to its existence and potential. Its first task was to conduct a rescue operation to seek out the most significant material created by research from past years. The second was to work with the ESRC to implement a Datasets Policy (ESRC 2002a) to ensure that for current and future projects the unnecessary waste of the past did not continue. Qualidata was not set up as an archive itself, but as a clearing house and an action unit, its role being to locate and evaluate research data, catalogue it, organise its transfer to suitable archives across the UK, publicise its existence to researchers and encourage re-use of the collections (Corti, Foster, & Thompson 1995, Thompson & Corti 1998).

Qualidata established procedures for sorting, processing and listing both raw data and accompanying documentation (meta-data); systematically describing studies for web-based resource discovery systems; establishing appropriate mechanisms for access; and promotion of and training in the re-use of qualitative data (Corti 2000). By 2002, Qualidata had acquired, processed and catalogued some 140 datasets, and catalogued a further 150 already housed in archives across the UK. Surviving "classic studies" data from key researchers were also rescued, including well-known British single projects such as Goldthorpe *et al.*'s *The Affluent Worker* (1962), Stan Cohen's *Folk Devils and Moral Panics* (1967), and the entire life's work of pioneering UK researchers such as Peter Townsend (*Family Life of Old People* (1955), *The Last Refuge* (1962) and *Poverty in the UK* (1979)) and Paul Thompson (the life-history interview studies of *The Edwardians* (1975) and *Families, Social Mobility and Ageing. An Intergenerational Approach* (1993)).

In the US, there is also a centre that has been systematically gathering qualitative research data in order to make it available to other social science researchers. Founded in 1976, the Murray Research Center: A Center for the Study of Lives is a national repository for social and behavioural science data on human development and social change, with special emphasis on the lives of American women (James & Sørenson 2000). The archive holds more than 270 data sets with a wide range of topics, samples, and designs. Many of these studies include in-depth interviews or, at the very least, some open-ended survey questions. One major collection of longitudinal studies of mental health includes Glueck & Glueck's *Crime Causation Study* (1968), The Institute of Human Development's *Intergenerational Studies*, and Terman's *Life Cycle Study Of Children Of High Ability* (1954). In the area of racial and ethnic diversity, an important study is Brunswick's *Harlem Longitudinal Study* (1994).

Finally, over the past few years there have been a number of other initiatives across the world that have sought to establish national archiving projects for qualitative research data. At the time of writing, the small-scale Czech Archive of Qualitative Data and Documents at the Faculty of Social Studies of Masaryk University has recently been established; Germany and Switzerland are currently preparing proposals for creating competence and archival resource centres for qualitative research; and others, led by national (survey-based) Social

Science Data Archives in Finland, the Netherlands, Denmark, and Canada, are conducting feasibility work.

A new era for qualitative data provision

From 2001, Qualidata began a new life as a specialist unit housed within the UK Data Archive (UKDA) at the University of Essex, with a focus on acquiring and distributing digital data. The key drivers behind merging the data services were the desire to create a one-stop social science data shop built around a single hub giving Essex a unique portfolio of data expertise and technological vision; the need to strengthen alliances to meet a tendering process ensuing from the ESRC's strategic review of its data archiving and dissemination services; the wish to streamline and simplify the data deposit process for ESRC depositors; and a growing need to reduce the demarcation between qualitative and quantitative data. It is also true to say that without the merger, the Qualidata service, which had suffered a significant cut in funding and loss of key staff during the review period from 1999–2001, would probably not have survived.

Phase I of the integration process was complete by October 2001, with many of the strategic and operational procedures for data acquisition, processing, meta-data creation and dissemination in place. Moreover, staff were fully integrated within the UKDA infrastructure. The period until December 2002 saw further efforts to harmonise working practices. First, a programme of cross-divisional training was initiated to broaden the data processing skills of UKDA staff to cover a wider range of data types, including mixed methods datasets. Second, the Qualidata web site and online catalogue were transferred to the UKDA servers. Finally, Qualidata has rolled out a programme of work to create freely available online User Guides for all its major collections.

New directions

There is a well-established tradition in social science of secondary analysis of quantitative data, and there is no logical intellectual reason why this should not be so for qualitative data. The research culture of re-using others' qualitative data is relatively young, and as such the body of published ''evidence'' about the benefits and limitations of the method is restricted. The build-up of stocks of qualitative data resources has encouraged the uptake of secondary analysis, but it is clear that the patterns of re-use of data witnessed by Qualidata since its inception in 1995 have been largely dependent on what data are on offer. As the stock of data grows, so the user base grows and experiences of secondary analysts find their way into the academic domain. That said, the demand for re-using data is partly a result of the efforts invested in repackaging and promoting data collections according to researchers' or learners' wishes, and following through with dedicated user support.

In recognition of the need to bring on board and engage new users of qualitative data resources, the ESRC embarked on an opening tendering process for a national qualitative data service, earmarked as a ''value-added'' specialist service of the greater ESRC/JICS Economic Data Service. The prime objectives of creating an ESDS integrated service were to provide the development and maintenance of a more integrated approach to data archiving and dissemination, and to provide more seamless and easier access to a range of disparate social science data resources for higher and further (post-16 non-compulsory) education.

The Qualidata unit thus has a new focus on providing access to, and support for, a range of user-friendly and accessible qualitative datasets. The work builds on Qualidata expertise and international reputation in this area, developed over the past eight years.

Central to the new service are a number of key objectives: the creation of a number of strategies for ''data enhancement''; a programme of work to improve access to data and documentation, for example, via the web; and finally to facilitate secondary usage through proactive user support and training activities. A six-fold strategy of data enhancement has been proposed comprising:

- the creation of *web-based samplers* to provide ''edited highlights'' of key qualitative materials to illustrate the potential of the collection for research and teaching;

- the creation of *thematic resources* whereby interviews relating to a particular theme and time period will be combined into a single resource, for example, crime and social order in the late twentieth century;
- *value-added processing*, ensuring that interview data are fully anonymised, are in an appropriate digital format, have speakers' tags assigned and have enhanced finding aids comprising dedicated online user guides, associated web pages;
- *web delivery* of marked-up primary data, such as interview transcripts, using XML standards and tools to facilitate rapid and flexible retrieval of information;
- *enhanced access* to key qualitative collections held elsewhere, in partnership with the hosting archive, to facilitate use in research and teaching;
- a *video-archive demonstrator* to investigate the use of video methods, focusing on methodological, ethical, technical, and analytical questions.

For Qualidata, under its new remit, a first step towards providing online access to data has been the development of the Edwardians Online project (Barker 2002), discussed below, an online resource that provides content-based access to a collection of oral history interviews with people who lived in Edwardian Britain. The multi-media resource integrates existing primary and secondary materials relating to the interviews, including the original text transcripts, digital sound-bites of the original audio tapes, background material concerning the original research study, and details of publications based on secondary studies of the interview texts. This resource has provided a model for the digitisation and interactive online provision of "classic collections" based on qualitative data for research and teaching resources.

The publishing of enhanced qualitative data resources within web-based systems will be supplemented by mounting all newly acquired core qualitative data on the UKDA's web-based direct download service. Previously acquired electronic data will be mounted on the download service according to user demand and data will typically be offered in word-processed or plain text formats rather than in a specific Computer Assisted Qualitative Data Analysis Software (CAQDAS) format (although the service will also meet any significant demand for software specific formats). One of the service's early priorities is to define and encourage the uptake of a software-independent data format both for longer-term preservation purposes, and for transporting coded data between CAQDAS packages.

For the new service, much emphasis is being placed on user support, including a dedicated help-desk facility, user events and training days, and "data confrontation" workshops to enhance the methodological and substantive understanding, and secondary analytical potential, of archived qualitative data sources. Advice and support will also be provided, as in the past, for creators and depositors of qualitative data. I return to support and training issues in greater detail further on.

Infrastructural needs for a dynamic new service

There are six areas relating directly to national infrastructure that are fundamental to enabling qualitative data provision, sharing and re-use in the ways users now desire and demand. These are probably best viewed as essential, rather than merely desirable, requirements:

- a high-quality national social science research base;
- a data archiving and dissemination infrastructure that is adequately funded, with foresight;
- mandatory data sharing policies;
- access to research and technical networks of expertise;
- access to a pool of "educated" and skilled users;
- centres of expertise, reputation and innovation.

A national social science research base

In the UK, the Economic and Social Research Council (ESRC) is the leading agency for funding research and training in social and economic issues, with a budget of over £78 million per year. It has an international reputation for providing high quality research on issues of importance to business, the public sector and government and a commitment to training excellence. The Council has a remit:

- to promote and support, by any means, high quality basic, strategic and applied research and related postgraduate training in the social sciences;
- to advance knowledge and provide trained social scientists who meet the needs of users and beneficiaries, thereby contributing to the economic competitiveness of the United Kingdom, the effectiveness of public services and policy, and the quality of life;
- to provide advice on, and disseminate knowledge and promote public understanding of the social sciences.

Its core strategic objectives are:

- to focus social sciences research on scientific and national priorities;
- to enhance the capacity for the highest quality in social science research;
- to increase the impact of ESRC's research on policy and practice;
- to deliver ESRC's activities effectively and efficiently.

Fortunately for the UK research communities, the ESRC appreciates the importance of preservation, sharing and re-use of social science data. There is already a recognition that research resources in the form of data created through research are long-term assets; that research must be of high quality, and problem-driven rather than specifically methods-driven; and that research, where appropriate, should be interdisciplinary and international in nature. I discuss the UK data sharing policy next, to highlight its vital role in the construction of a national stock of qualitative data resources.

National data archiving and dissemination infrastructure

In addition to key national research funders proclaiming their support for research resource provision and a culture of data sharing and secondary usage, a national infrastructure must be in place in order to back up the words with action. Ideally, this would accommodate:

- centres of expertise;
- skilled leaders and highly trained staff;
- in-house data processing and preservation activities, for which the costs are adequately recognised;
- the need for technical developments;
- user support and training activities – from both the reactive and proactive points of view;
- a programme of research to undertake both value-added and methodological work on data.

The UK is comparatively fortunate in having seen the ESRC channel extensive resources into a national data archive since 1967, and smaller sums of money into a qualitative data service since 1994. The good fortune is set to continue for the foreseeable future, with guaranteed funding until 2007, covering many of the elements I have listed above. However, the robustness of these centres is somewhat dependent on the ESRC having championed and supported a formalised Data Sharing Policy for some years. In the next section I provide an overview of the ESRC Datsets Policy, because, certainly for Qualidata, in many ways it is the lynch pin. Indeed, for Qualidata, its data acquisitions strategy is largely dependent on the inflow of data arising from researchers' obligations to adhere to the Policy.

Mandatory data sharing policies

There are a number of key drivers for establishing data sharing policies. The principle issue is the growing perception that "data" are the primary building blocks of science. Second, legal requirements and public funding arguments are convincing motivations for research funders to establish mechanisms for enabling access to data. Third, demand from the research communities to gain access to expensive already-collected data and the willingness to share their own data helps to get the issue onto the policy agenda. Finally, dramatic advances in the conduct of scientific research that collects massive amounts of data, which are often distributed and require expensive storage and analysis facilities, require suitable infrastructures to be in place. Opposing the drivers are barriers that can complicate data-sharing policies – those of property rights and public privacy – although neither of these is insurmountable.

Enabling meaningful access to reliable scientific data merits attention to its preservation, archiving, and sharing. Many funding bodies now recognise that there are a number of persuasive reasons for investing in data sharing. The National Institute of Health summarises these in a concise way: "sharing data reinforces open scientific inquiry, encourages diversity of analysis and opinion, promotes new research, makes possible the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and measurement, facilitates the education of new researchers, enables the exploration of topics not envisioned by the initial investigators, and permits the creation of new data sets when data from multiple sources are combined. By avoiding the duplication of expensive data collection activities, the NIH is able to support more investigators than it could if similar data had to be collected afresh *de novo* by each applicant …".

However, investment in data sharing is as yet uneven across disciplines. The social sciences and humanities have led the way in implementing and promoting data policies, in some cases boasting a 30-year investment profile. Consideration of funders' policies underlines this. In the UK, while many research funders do operate data-sharing policies, guiding principles are most evident in the social sciences and humanities. Among the natural sciences only the NERC has a formal data policy. These policies vary in how mandatory they are; how involved the recipient organisations are in appraising research applications and associated data management plans; the degree to which in a budget line should be costed in for data preparation and documentation for archiving; and in rules on allowing researchers' to place embargoes upon data.

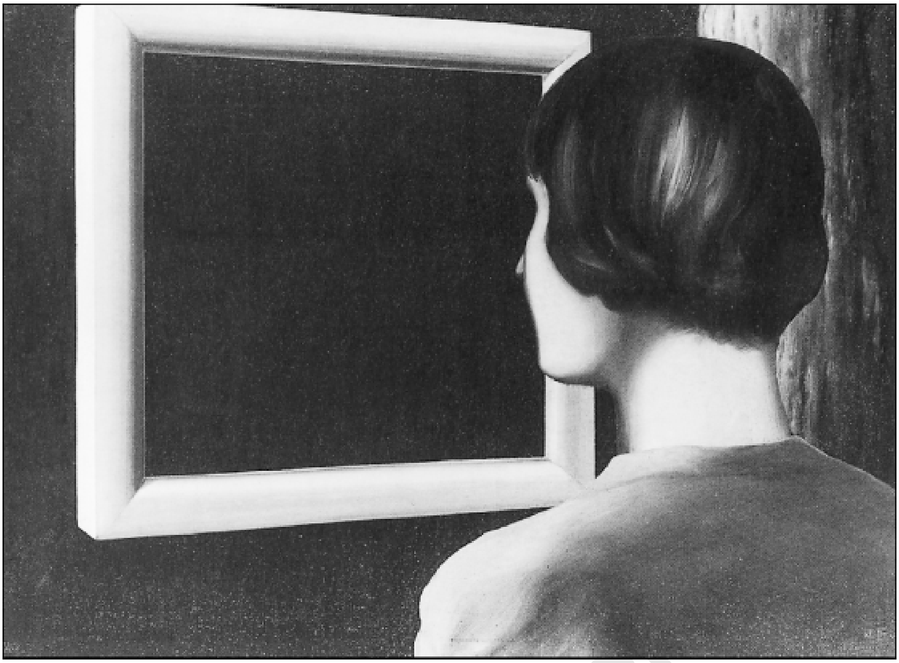
The ESRC Datasets Policy was established in the 1990s and reinforces and emphasises the ESRC's stated position relating to the acquisition and use of datasets, the requirements of which are now a condition of ESRC research funding. The ESRC requires all award-holders to offer for deposit copies of both machine-readable and non-machine-readable qualitative data within three months of the end of the award. This relates not only to datasets arising as a result of primary data collection, but also to derived datasets resulting from ESRC-funded work.

In order to operate the Datasets Policy, the ESRC supports two Resource Centres with responsibilities for the cataloguing and archiving of data. The UKDA is responsible for acquiring, documenting, disseminating, and preserving digital data created during the course of ESRC research grants. Qualidata has special responsibility for qualitative data in both digital and non-digital form. The UKDA and Qualidata have a coordinated quantitative/qualitative acquisitions strategy that encourages the stream of qualitative data destined for archiving. Both centres have long-standing experience dealing with all aspects of acquisition and data collections management, including licensing agreements, working with academic award holders in the process of depositing data, and established relationships with other data producers, such as other research funders, and are well placed to operate the Policy.

The Policy requires that datasets must be deposited to a standard which would enable the data to be used by a third party, including the provision of adequate documentation. Depositors are advised to contact the two Resource Centres at the earliest opportunity should the nature of the data be such that it may be difficult to lodge. The earlier in the research process these discussions occur, the more likely researchers are to create datasets that are well-documented, free of confidentiality or licence constraints, and useable for secondary analysis. Support for award holders and potential depositors is generally provided through web-based guidelines and notes on preparing data for deposit. Support extends to adopting a more proactive role, working to promote the importance of sharing and preserving data within the social sciences, and actively alerting award holders to their obligations.

Copyright in data deposited with the UKDA is retained by the copyright-holder(s), with whom terms for access to the data are agreed. Deposits are to be accompanied by a signed licence form. Use of data is also subject to acceptance by the user of a formalised access agreement complying with the terms and conditions of deposit.

Close collaboration between data specialists, data creators, and users in the development



L'Image parfaite [The perfect picture] (1928), by René Magritte (1898–1967). Private collection. ADAGP.DR.

of data-sharing standards, tools, and infrastructure, is paramount, particularly with respect to the following matters:

- meta-data and associated standards;
- data custodianship and rights management framework;
- consent for future "new use" of data;
- data quality assurance;
- preservation standards and tools;
- criteria for prioritising resources for investment in preservation and sharing.

The UKDA and Qualidata were instrumental in helping to draft and establish the ESRC Datasets Policy and, based on their experience of "operating" it over the last 10 years, have recently proposed a set of changes to the operational procedures that should create a more robust, systematic and accountable Policy. One of the central concerns of the current policy is that improved three-way communication channels between the ESRC, the award holders and the data archiving and dissemination services would be highly beneficial to the Resource Centres. The first suggestion was that the archiving and dissemination services should be involved across the life-cycle of data generation, in particular enabling Resource Centre to have input at the grant application selection stage to encourage deposit of high-quality data and documentation. Second, in order to put the first proposal into place, the ESRC need to establish a fully coordinated strategy in-house, with dedicated staffing, to ensure the smooth running and auditing of the policy. An example would be for the UK archiving centres to receive timely updates about new grants and data creation activities, which at present does not happen.

Third, the Datasets Policy would, like the NERC principles outlined above, benefit from a requirement for data creators to produce a formalised data management plan at the application or short-listing stage, particularly for expensive research programmes; a more stringent view towards the length of time allowed for data embargo demanded by investigators; and enforcement of penalties for "non-compliant" researchers. The ESRC Datasets Policy is currently under review but a summary of the recent Policy can be found in section 17 of the ESRC Guide to Research Funding (ESRC 2002a).

Finally, it would also be highly beneficial to data services if four other strands of research-related activity were incorporated into the data-sharing policy and associated with peer review of research grant applications:

- peer reviewers should advise on the long-term value of research data and the wider community should recognise high-quality data as a valued research output;
- an education programme is required to induce a shift in attitudes towards informed consent, confidentiality and copyright of data and recognition of an extended lifetime for data by participants, investigators, funders (academic and non-academic), research ethics committees, and policy and law makers;
- research programmes should look at their within-project data generation and data documentation activities in a more complementary and systemic way; and
- there should be specific calls for research grants to conduct secondary analysis of archived data.

Access to research and technical expertise: developing data resources

Online provision of data resources can require significant research and development activities that are costly and require concerted evaluation, usability testing, and roll-out time. Equally, even when adopting technical solutions from other areas of development, adapting the product to suit the organisation's own technical needs can be just as daunting and time-consuming. The UKDA has typically not received support from its core funders for any kind of research and technical developments, and has relied on drawing in grants from other sources. For example, NESSTAR (Networking Social Science Tools and Resources) which was a multi-country project, attracted some £2 million from the European Commission under the 4th and 5th Information Technology (IT) Framework Programmes to produce a suite of online survey data-browsing and exploration tools. The LIMBER project (Language Independent Metadata Browsing of European Resources), based at the UKDA, also attracted funding from the European Commission to create a multilingual user interface to the data stored in social science archives across Europe.

Maintaining a vibrant data community, such as through the Council of European Social Science Data Archives (CESSDA), the International Federation of Data Organisations for the Social Sciences (IFDO) and the International Association for Social Science Information Service and Technology (IASSIST), is critical for allowing these project-oriented partnerships to emerge. The UK Joint Information Systems Committee, which is a co-funder of the new national Economic and Social Data Service, also frequently has calls for IT pilot or demonstrator projects that can often accommodate social science applications.

Our experience from the UK tells us that it is highly productive to forge links with funders in information and communication technologies and with developers who are addressing applications in disciplines outside of the social sciences. For example, the UKDA has a History Data Service in-house that is hooked into programmes that are developing digitised humanities resources.

Finally, in order to ensure that we are expending our efforts in the right areas, we must consult with our user and potential user communities on what they want, and how they want it. They must be involved in all phases of research and development work, be it consultation, testing, or evaluation.

Access to a skilled user base

If secondary analysis of qualitative data is to become a commonplace and accepted method for the social sciences then we need a body of instructional literature on the methods, together with published case studies that are seen as exemplary. Essentially, a new "breed" of skilled users must be grown and nurtured, to whom we first need to "feed" comforting and easily digestible learning materials that demonstrate appropriate methodological and analytic strategies.

A number of resources or activities, which perhaps seem a little idealistic, that may help facilitate secondary analysis of archived data by novice and experienced researchers alike:

- effective resource discovery tools and useful meta-data;
- informative user-oriented web sites with advice, downloadable materials, case studies, “frequently asked questions” (FAQs), etc.;
- publications based on secondary analysis of qualitative data recognised in the literature;
- a long-term training programme in “re-using” data;
- access to a broader national training programme in research skills (such as research design, data collection, analysis, and write-up);
- active user communities across the domains of research, teaching, and learning that are both harnessed and self-supporting;
- interdisciplinary and international working relationships and partnerships.

I will focus here on training needs, as in many ways they are the key to encouraging the uptake of secondary analysis in appropriate ways.

Plugging into the learning and teaching communities

Archived qualitative data are a rich, unique yet often unexploited source of research information for teaching and learning. Whilst the culture of sharing and re-using has become far more widely accepted in the UK, largely promoted by Qualidata, surveys suggest that specific training resources on re-using data are sought after and would be welcomed. It is thus unfortunate that provision for these communities was explicitly excluded from Qualidata’s remit by its funders. In spite of this, Qualidata user-support staff have always been highly receptive towards approaches from users who require data for teaching, and have prepared specialist sets of interviews for teaching on a variety of courses: introductions to CAQDAS packages, oral history, discourse analysis, and general research-methods courses.

Qualidata hosts web pages on using data in teaching and learning. While published information can help students to confront data, students are demanding users. Many of the queries tracked by both Qualidata and the UKDA could be, potentially, highly resource intensive in terms of staff time if they were answered in full. For example, many postgraduates ask very specific questions, which often reflect the title of their thesis, for example, “what analyses would I have to undertake to measure gender inequalities in health?”. At best, support officers can direct them to relevant sources of data and suggest types of analytic strategies, but they are in fact briefed to refer demanding students back to their tutors – or to advise them to sign up for training.

Great emphasis in the new Qualidata Data Service will be placed on user support, and a dedicated help-desk facility will be established. Tailored user guides, theme-based web pages, and FAQs will be produced. User events and training days will be organised in line with identified user needs. Workshops will include generalist introductory sessions and more focused meetings on detailed areas of research interest and methodology. These will be supplemented by “data confrontation” workshops to enhance the methodological and substantive understanding, and secondary analytical potential, of archived qualitative data sources.

The main thrust here is to provide students with the opportunity to learn about many fundamental aspects of qualitative research in addition to gaining first-hand experience of reanalysing, comparing, and critiquing data from a variety of sources. Indeed the concept of reusing of data becomes tangible once time is spent examining data-rich collections. An appreciation of research methods employed in “classic studies” can also be better grasped when examining the contextual information about the study, such as topic guides, field notes, analytic notes and resulting published and unpublished reports. Learning about the work of researchers who have made a significant impact in their field allows young researchers to take the best practice elements from this work and further develop them in their own research work. Moreover, by illustrating the importance of planning data collection and management with future re-use in mind, they may be more inclined to archive and share their data further down the track, and also to think imaginatively about re-using their own data.

Creating and delivering more visible and packaged online electronic resources is a key way to facilitate both the usage of data and training in methodological skills among students. In order for these products to be of most benefit, they need to be accompanied by: substantive and methodological commentary on the project and data, hands-on exercises; the availability of face-to-face training; and finally follow-through individual support. In order to pool expertise and maximise the use of available resources such deliverables are best achieved via collaborative initiatives. The UKDA already has experience in this task in relation to both Training Resources and Materials for the Social Sciences (TRaMSS) and the Resource Discovery Network (RDN), and seeks to build closer links with other training initiatives in the social sciences, such the recent ESRC Research Methods Programme (see below). Joint events are also planned with other national service providers and training initiatives such as the CAQDAS networking project at Surrey.

In 1996, Qualidata prepared a teaching pack based on the *Edwardians* data (Thompson 1975), which described oral history methods and presented ways of re-using this data collection. The pack was well regarded and widely used in teaching. The new Edwardians Online resource will build on this concept and will feature freely available associated training exercises geared towards a wider range of educational levels.

In short, both reactive and proactive support for student learning is vital for maintaining user numbers. In terms of promotional and support strategies that may help to facilitate usage of data in these domains, I would recommend:

– targeting key departments/relevant discussion lists with promotional and training materials;
– offering/agreeing to talk to postgraduate students locally and in other key social science departments across the country;
– liasing with local and national learning and teaching organisations;
– publishing in teaching and graduate outlets;
– seeking specific grants to produce dedicated teaching and learning materials; and
– encouraging teachers to get involved in evaluating training resources.

Broader research-methods training requirements

The UK, like many other countries where there exists a flourishing social science research community, has suffered from the lack of any joined-up strategy for training in conducting research and data analysis. It has been impossible even to identify, in a coherent way, existing training programmes available to researchers, students, and professionals.

However, there have been a host of new strategic initiatives that have begun to address this major deficit. These are the ESRC's Research Methods Programme (ESRC 2002b), new Post-Graduate Training Guidelines (ESRC 2002c), and new Research Resources Board Strategy (ESRC 2002d).

Phase I of the Research Methods Programme, for which awards were granted in the spring of 2002, saw relatively small-scale methodologically oriented projects, while the forthcoming Phase II is set to concentrate on supporting national quantitative and qualitative research training provision. A recent consultation meeting on training hosted by the Programme concluded that: "Training needs to be ongoing for researchers at all levels – from graduate students through to senior researchers. It was emphasised that developments in methods require constant up dating. It is important that the trainers receive training to ensure that up-to-date skills are passed on to students – however this may not be easy to achieve…(and) Training needs to be closely linked with substantive research questions and, generally, needs to be disciplinary based. Interdisciplinary training is valuable but needs to build from and across the different disciplinary/substantive bases, rather than adopting a purely generic approach." (Dale 2002).

The meeting's overarching sentiment was that a joined-up strategy was vital for enabling future generations of skilled researchers and competent data analysts.

In 2001–2, the ESRC also commissioned a review of its Post-Graduate Training Guidelines, with key input from all the social science disciplines. The upshot of the consultation was to formulate a national strategy to ensure high-quality relevant training in methods and analytical skills for ESRC funded postgraduate students. Subsequently, new Guidelines were produced: "to indicate the skills and competencies that postgraduate research students should have acquired by the time they have completed a research degree, if they are to be accepted as professionally trained researchers in their subject; (…) to outline broadly the overall context, objectives and content of the training that students must have received by the time they have completed a research degree; (…) to provide criteria for ESRC's assessment of master's courses and doctoral provision. Fulfilment of these criteria allows successful applicants for ESRC recognition to receive ESRC studentships.'' (ESRC 2002c).

Finally, the ESRC Research Resources Board, under which the ESDS and Qualidata are funded, has revamped its longer-term strategy to take on board the provision of high-quality resources and training requirements for researchers. The Resources Board's purpose is to: ''support Council policies through the provision of research resources for the social sciences and advise the Council on the provision needed to ensure the long-term vigour and utility of the social sciences and for high quality research.'' Specifically, ''The 'research resources' funded by ESRC can include source material such as local, national and international qualitative, quantitative and spatial data; the housing, maintenance and provision of access to this information in archives or resource centres; library holdings; software; communications technologies and other hardware. To exploit these resources, the Board will (ensure that) research is maintained but also improved both to meet the needs of users and stakeholders. The next generation of researchers are trained but also include updating opportunities for experienced researchers. Central to achieving this is the reskilling and enhancing skills of those engaged in training where necessary.'' (ESRC 2002d).

International training and mobility opportunities

Cross-national training programmes aimed at training in secondary analysis are few and far between, and those covering the qualitative research tradition are rarely to be found. Exceptions include various summer schools, such as the Essex and Swiss Social Science Data Analysis summer schools, which do support qualitative research and methods training. Under the European Union Large-Scale Facilities Activity, the European Centre for Analysis in the Social Sciences (ECASS), an interdisciplinary research centre at the University of Essex, also provides the support services, usually in the form of short-term placements, for researchers to work on archived data on location.

Additionally, there is limited funding available for researchers wishing to travel to another European country to undertake a small piece of research, which could, feasibly, include a project utilising archived qualitative data. Examples are EC Marie Curie Short Stay Fellowships for Doctoral students and European Consortium of Sociological Research (ECSR) small grants for exchange of research and post-doctoral students. The research communities would benefit from funders supporting a greater number and range of these kinds of activities.

Centres of expertise: reputation and innovation

Success for a qualitative data service is also dependent on having demonstrated strengths in leadership, management, and vision. These include having:

- respect within the national academic qualitative research community;
- local institutional standing and support;
- an excellent governance framework;
- good links with the data-resource oriented technical community;
- productive with the teaching and learning communities; and
- recognition by the wider international data archiving community.

In particular I would single out the need for a high-quality user support team and programme of support work pulled together by coherent and forward-thinking direction and management that keeps one step ahead of users' needs. Support activities, whether reactive or proactive, raise the profile of a research resource-based organisation. High-class support pays off in terms of yielding a good reputation; a solid funding base; an enhanced culture of sharing and re-using qualitative data; the production of high quality incoming data and documentation; and spin-off funding for new products and cross-national initiatives.

In many respects, the UKDA and Qualidata are often hailed as pioneers. However, I do believe that the "competence centre" model, as currently proposed for qualitative research and data in Switzerland and Germany, may hold the key to a new future for the secondary analysis of qualitative data. This model combines provision and support for archived data with an active in-house programme of research and individually tailored support for researchers, and probably offers the best opportunities for users. With the right funding, structure, direction and management, staffing component, and partnerships with related areas of national research expertise, I would hope that these centres might prove exemplary in the re-use of qualitative data.

Increasing access to qualitative data: online digital resources

The issue of how to make these data resources accessible to users is a central concern for Qualidata, which is 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. The service has also initiated an increased focus on depositing digital data in-house with the UKDA, and the digitisation of "classic collections" for research and teaching resources. Access to qualitative data has been facilitated by mounting on the UKDA's instant web-based download system. Using such services, registered users can acquire digital data collections at the click of a mouse, rather than making visits to special collections to spend time working through boxes of paper transcripts.

In early efforts in this content-oriented direction, in response to user demand, the emphasis has been on data development, with a view to providing users with direct access to the content and structure of digitised collections via an online facility. This can be viewed as a significant step beyond file download, whereby a user can download a set of interview transcripts and import them into a data management software package. When we talk about content and structure in context of this development work, we are concentrating on features such as speaker tags, coded data, and links to contextual material (audio, fieldnotes, photos, analytical annotations, etc.).

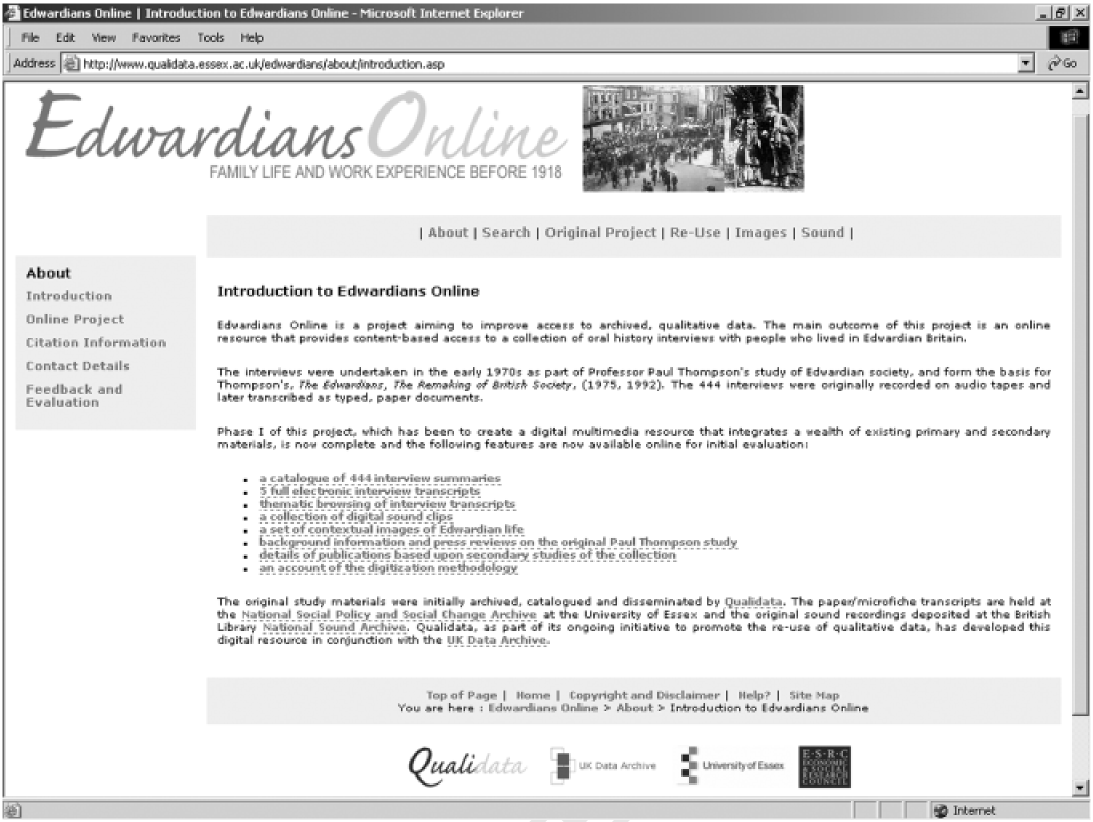
The Edwardians Online data collection

Edwardians Online, which is based upon a set of oral history interviews, was selected as an appropriate collection for undertaking Qualidata's first major web-based digitisation project.

The interviews under consideration were undertaken in the early 1970s as part of Professor Paul Thompson's study of Edwardian society, and form the basis for his *The Edwardians. The Remaking of British Society* (1975). The 444 interviews drawn from a cross-national sample of people born in Britain before 1918 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. In spite of the non-digital format of the interviews, the paper source has proved to be very popular for re-use. Indeed, the collection has attracted high usage across a variety of research interests and has value as a teaching resource. Users have requested access to both complete interview transcripts and more specific information or extracts from within the documents. Because the interviews are long, using 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. Moreover, data exist in various formats in various locations: originally recorded on audio tapes; transcribed as typed paper documents; text extracts coded and pasted in paper form during the thematic analysis of the content; supporting source materials, such as essays and letters. Finally, the data are representative of a broad class of qualitative interview data.

In June 2002, Qualidata released Edwardians Online, a pilot web-based, multimedia, digital resource. The aim of this work is to develop a standard framework and a demonstrator for providing online access to the content of digitised qualitative data collections. The pilot resource integrates a wealth of existing primary and secondary materials from the oral history study. A database of the interview summaries and a sample of full text transcripts can be searched using free text or by theme, the


Edwardians Online | Introduction to Edwardians Online - Microsoft Internet Explorer
File Edit View Favorites Tools Help
Address
Go
Edwardians Online
FAMILY LIFE AND WORK EXPERIENCE BEFORE 1918
Qualidata
UK Data Archive
University of Essex
E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL
Internet

http://www.qualidata.essex.ac.uk/edwardians/about/introduction.asp

| About | Search | Original Project | Re-Use | Images | Sound |

About
Introduction
Online Project
Citation Information
Contact Details
Feedback and Evaluation

Introduction to Edwardians Online

Edwardians Online is a project aiming to improve access to archived, qualitative data. The main outcome of this project is an online resource that provides content-based access to a collection of oral history interviews with people who lived in Edwardian Britain.

The interviews were undertaken in the early 1970s as part of Professor Paul Thompson's study of Edwardian society, and 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.

Phase I of this project, which has been to create a digital multimedia resource that integrates a wealth of existing primary and secondary materials, is now complete and the following features are now available online for initial evaluation:

- a catalogue of 444 interview summaries
- 5 full electronic interview transcripts
- thematic browsing of interview transcripts
- a collection of digital sound clips
- a set of contextual images of Edwardian life
- background information and press reviews on the original Paul Thompson study
- details of publications based upon secondary studies of the collection
- an account of the digitization methodology

The original study materials were initially archived, catalogued and disseminated by Qualidata. The paper/microfiche transcripts are held at the National Social Policy and Social Change Archive at the University of Essex and the original sound recordings deposited at the British Library National Sound Archive. Qualidata, as part of its ongoing initiative to promote the re-use of qualitative data, has developed this digital resource in conjunction with the UK Data Archive.

Top of Page | Home | Copyright and Disclaimer | Help? | Site Map
You are here : Edwardians Online > About > Introduction to Edwardians Online

FIGURE 1.

latter 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 is also included. The second phase of the project plans an extension of these features, such as linking to other key sources such as maps and census data from the period.

Phase I of the project

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

- the problem of developing a non-proprietary electronic format for preserving the content of qualitative datasets;
- the development of tools for facilitating the encoding of data in this format; and
- the question of the methods of access and facilities for exploring qualitative data online.

An important early lesson for the project was the need to look outside the data archiving and social science communities, in which development work in this vein simply cannot be found. The project staff thus drew on humanities scholars for their experience in creating (and receiving huge grants for) web-based text and digital resources; computer scientists for XML for data storage, manipulation, and web-presentation tools; and computational linguistics researchers for their expertise in natural language processing and information extraction.

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:

- support the encoding of the content of various types of primary data documents produced in qualitative research;
- support the encoding of contextual documentation and meta-data linked to the primary sources;
- be capable of providing formalised links between the texts and associated audio and video materials, with a view to providing in the long term, integrated, multimedia resources;
- be able to 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 in that 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).

The DDI provides an XML framework for study descriptions of social science datasets, yet it cannot represent the content of qualitative data as it can for survey data (e.g., browsing variable frequencies online).

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:

- the types of documents and structures to be marked up;
- formal definition of a common XML vocabulary and Document Type Definition (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, and research to date has considered two options. The first was to create a customised application of XML specifically for the purpose of marking up the content of spoken interviews and other types of qualitative material. The second is 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.

Phase II

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

Conclusion

Qualidata has a new future that has been enabled by a medium-term national strategy with a renewed emphasis on access to user-friendly data and complementary support services. The five-year UK strategy comes in the form of a new partnership between two major funding organisations: the ESRC and the JISC. Three years ago, these two funders might not have considered jointly funding a national data service for the social sciences. Early days of contract negotiation have already highlighted the disparate, and sometimes opposing notions, expectations, and service level requirements envisaged by the two funders. The fact that the new Economic and Social Data Service has kicked off in a relatively smooth manner suggests that the partnership is working, which can be largely attributed to the complementary strengths of the partner organisations and the synergy that has been built between them at many levels, including policy, staffing, and activity.

I wish to conclude by summarising what I consider to be the key infrastructure elements for running a successful national service for qualitative data. First, there needs to be national recognition of the long-term value of data resources, alongside the traditional intellectual products of research, the investment of which are managed under a formalised Data Policy. Second, a Datasets Policy must respect the range of types and formats of data created in the course of social and economic research, and must set in place an appropriate legal and ethical framework that will enable more sensitive data to be accessed. Third, users' needs must be supported and reviewed, and the knowledge acquired by support staff fed back into study documentation and theme/analysis-related information on an iterative basis. Fourth, outreach and training programmes must be geared to meet users' skills levels and requirements, and with a greater motivation in mind, join forces with other training providers to help to redress the skills deficit in research design and conduct and data analysis.

Finally, providing access to qualitative data is, in part, dependent upon new technologies. The recent developments for XML standards and tools for data storage and retrieval, which enable us to build resources such as Edwardians Online, did not exist 10 years ago. Equally, we will undoubtedly see many of the data preparation tasks for qualitative data that are so labour intensive (such as manual indexing of audio material or routine anonymisation of data) become simple "click of a button" tasks using hidden and highly intelligent software. These may be dreams, but they are bound to become realities, and in this respect, Qualidata must look to forming interdisciplinary partnerships with researchers outside of its own domain, in particular language-processing experts and engineers.

References

BARKER, E. 2002. *Edwardians Online Pilot Resource*. Qualidata/ UKDA.

BRUNSWICK, A. 1994. *Harlem Longitudinal Study of Urban Black Youth 1989–1994*. [dataset archived online at http://www.radcliffe.edu/murray/data/ds/ds0845.htm].

CORTI, L. 2000. "Progress and problems of preserving and providing access to qualitative data for social research. The international picture of an emerging culture". *Forum Qualitative Social Research*, 1(3). [available online at http://www.qualitative-research.net/fqs-texte/3-00/3-00corti-e.htm].

CORTI, L., FOSTER, J., AND THOMPSON, P. 1995. "Archiving Q3qualitative research data". *Social Research Update*, 10.

CORTI, L., AND THOMPSON, P. 2000. *Annual Report of Qualidata to the ESRC*. University of Essex.

DALE, A. 2002. "Research Methods Programme Consultation meeting on training". *a summary of key points*, November 29 2002. Document available online at http://www.ccsr.ac.uk/methods/archive/consultationmeeting/keypoints.shtml.

ESRC ECONOMIC AND SOCIAL RESEARCH COUNCIL 2002a. *ESRC Datasets Policy*. Swindon: ESRC.

ESRC ECONOMIC AND SOCIAL RESEARCH COUNCIL 2002b. *ESRC Research Methods Programme*. Swindon: ESRC [available online at http://www.ccsr.ac.uk/methods/].

ESRC ECONOMIC AND SOCIAL RESEARCH COUNCIL 2002c. *Postgraduate Training Guidelines 2001*. Swindon: ESRC [available online at bhttp://www.esrc.ac.uk/esrccontent/postgradfunding/postgraduate_training_guidelines_2001.asp].

ESRC ECONOMIC AND SOCIAL RESEARCH COUNCIL 2002d. *The ESRC Research Resources Board's Strategy for Supporting Research in the Social Sciences*. Swindon: ESRC [available online at http://www.esrc.ac.uk/esrccontent/aboutesrc/rrbstrat.asp].

GLUECK, S., AND GLUECK, E. 1968. *Delinquents and Nondelinquents in Perspective*. Cambridge, MA: Harvard University Press.

JAMES, J. B., AND S øRENSEN, A. 2000. "Archiving longitudinal data for future research. Why qualitative data add to a study's usefulness". *Forum Qualitative Social Research*, 1(3). [available online at http://www.qualitative-research.net/fqs-texte/3-00/3-00jamessorensen-e.htm].

REX, J., AND MOORE, R. 1967. *Race, Community and Conflict*. Oxford: Oxford University Press.

STACEY, M. 1974. "The myth of community studies", in Bell, C., and Newby, H. (Eds), *The Sociology of Community*. London: Frank Cass.

TERMAN, L. M. 1954. "Scientists and nonscientists in a group of 800 gifted men". *Psychological Monographs: General and Applied*, 68(7), 1–44.

THOMPSON, P. 1975. *The Edwardians. The Remaking of British Society*. London: Granada.

THOMPSON, P. 1991. *Pilot Study of Archiving Qualitative Data: Report to ESRC*. Department of Sociology, University of Essex.

THOMPSON, P., AND CORTI, L. 1998. "Are you sitting on your qualitative data? Qualidata's mission". *Social Research Methodology: Theory and Practice*, 1(1), 85–90.