

Qualitative research and E-Social Science: appraising the potential

Table of Contents

(c) Nigel G. Fielding

NB Software products and proprietary systems are shown in this report in italics and it is to be understood thereby that they are protected by relevant copyright and trademark law.

Executive Summary

With the national e-Science agenda advancing rapidly, and a growing ESRC investment in E-Social Science, it is timely to assess the potential returns to *qualitative* methodology and research from Grid and HPC technologies. UK social science has particular strengths in qualitative methodology, and qualitative research is increasingly important in policy-related, applied and private sector social research as well as in academic work. E-Social Science has the potential to enhance the scope, depth and rigour of qualitative inquiry. It can raise technical capacity and facilitate key operations. The present report scans existing applications of E-Social Science to qualitative research, indicates potential applications, and identifies promising areas for feasibility studies.

The broad potential of E-Social Science is captured by the threefold distinction between the Computational Grid (supporting high performance computing, HPC), the Access Grid (for collaborative work amongst dispersed researchers) and the Data Grid (for moving data between sites of storage and use). Large gains in computing resource enable new data-handling capacity and analytic procedures, new opportunities for work with colleagues who are not co-present, and new facilities to archive, curate and exploit the many kinds of data from which qualitative researchers make sense of social phenomena. From these capacities also flow new challenges to conventions of privacy, data integrity, and the relations that scientific communities have with society. These too are objects of study for qualitative E-Social Science, and for branches of the arts and humanities that engage with the ethics of contemporary society. Indeed, affinities in modes of inquiry between qualitative social research and research in the arts and humanities make for several connections between the interests of these fields in their encounter with e-Science.

The report's appraisal of qualitative E-Social Science draws on research conducted in July-August 2003, involving an e-mail questionnaire sent to UK and international qualitative methodologists and computer scientists, follow-up sub-sample interviews, group discussions with qualitative researchers, and response by the AHRB e-Science committee to a document outlining qualitative E-Social Science. The questionnaire, interviews and group discussions asked participants to outline their current engagement with E-Social Science, identify potential applications of E-Social Science that could make a difference in their particular field, and comment on aspects of E-Social Science, e-Science or the social impact of a pervasive computing environment more widely, including ethical matters. Current and potential applications were mostly elaborated in generic terms but some respondents identified applications to particular substantive fields. As >early adopters= of new technologies for social research, questionnaire and interview respondents were experienced in applying IT in qualitative research and alert to the potential of E-Social Science. The report's reflections on the potential of E-Social Science are, however, balanced by appreciation of the perspective expressed in the group discussions by those for whom E-Social Science comes as a new thing.

The existing qualitative applications of E-Social Science

Most respondents were already working on an E-Social Science activity. One concentration of work was on database development, including a database for generating hypotheses in studying linguistic morphology, one offering research instruments and other tools for online data collection, and another involving tools for content analysis of large natural language corpora. In the qualitative software (CAQDAS) field, E-Social Science related to current work on methodological interrelation, extending CAQDAS to analysis of visual data, and applying AI routines for more systematic qualitative data analysis. An adjacent interest was in applying XML to support CAQDAS-type tools for Web use, which would enable co-working by dispersed researchers and new opportunities for research subjects to participate in interpreting findings. Another concentration of work was on technologies to support ethnographic fieldwork, including interactive collection of data and the accessing and aggregation of data (and analytical tools) from field sites with that held at the fieldworker's home base. This involved developing XML-based Grid resources, middleware, and the use of a >micro-Grid= of GPS-enabled personal digital devices in the field. Another interest is in using PDAs to exchange information among social agencies, such as those relating to young offenders, with qualitative databases and GIS being used in real time to identify socio-economic and geographical patterns in risk factors.

Work with streaming video data is prominent, with a chief application in workplace interaction. This area involved extensive collaboration between dispersed research teams before the advent of E-Social Science, so Access Grid potential was quickly realised. Current effort is on requirement specification for real-time distributed manipulation, annotation and analysis of streamed data. An interest relevant to methodological integration is in >scaling up= findings from small-scale studies so that inquiries by cognate qualitative researchers build on each other. An XML metadata model shows promise, alongside protocols to standardise and link conceptual and substantive dimensions of individual studies. There is experience in using multimedia hypertext ethnography as a testbed for an integrated environment for the research process from data collection through analysis to authoring and dissemination, and in collecting data from online communities. There is no UK equivalent of the *qualitative-research.net* online gateway of the German universities but the initiative crystallises several pedagogical and professional affordances of E-Social Science, including

mechanisms to enhance online tutoring and peer review of professional work by exploiting the technical capacities of e-publishing. A final significant interest is in understanding the implications of E-Social Science for research ethics, data protection and regulation, an interest that also leads to wider appraisal of the prospects for a pervasive computing environment.

Current qualitative applications of E-Social Science thus include: archiving/database work; text analysis and content analysis; Grid technologies for fieldwork; visual methods (prominently, streaming video); Access Grid Nodes as sites of co-working; teaching and publishing media; ethics/data protection/regulation. On the basis of profiling current engagement with E-Social Science, and forecasts of developments by those involved, it is possible to say that E-Social Science has the capacity to change how qualitative researchers work, the kinds of substantive problems to which qualitative research is applied, and the integration of qualitative research findings with those of other methodologies.

The potential qualitative applications of E-Social Science

If one polled IT-oriented qualitative researchers on the E-Social Science applications of most immediate interest, the front runner would be remote access to databases, followed by work with streaming video, qualitative text analysis, and strong showings by Access Grid Nodes (AGNs), applications of E-Social Science to fieldwork, and automated content analysis. About three quarters of respondents were interested in ethical and information management aspects of the Grid but few saw it as their main interest in E-Social Science. About half identified problems in their field that could not presently be addressed due to constraints on computing power, and similar numbers identified other technological bottlenecks, including transcription applications, import/export facilities in CAQDAS to get data onto the Web, metadata systems, secure means for e-publishing video data, OCR systems for humanities data, and accommodating access/use rights of arts and humanities materials. But less interest lies in >headlines= than specific applications, to which we now turn.

Archival qualitative data remains under-exploited and qualitative secondary analysis is a limited practice. But interest in collaborative archival research is growing, and *applying E-Social Science technologies to archival databases* promotes it. Data Grids enable researchers to create and share annotations and there is keen interest in accessing dispersed, multi-modal distributed archival material with a view to producing multiple, inter-linked analytic narratives. Such applications call on multiple tools (for sound, image, transcript) and pose technical challenges including access control, concurrent update and integrity rollback. But such applications interest several disciplines, including anthropologists working with visual archives, linguists with sound archives, and humanities researchers, as well as qualitative social researchers. Grid resources can support novel data-sharing modes that promote re-use. *Access Grid Nodes* are presently often understood as enhanced video-teleconferencing. Exploiting them as a collaborative mode for qualitative research is an application waiting to happen. But this is only an extension of the >e-meeting= mental set. While AGNs provide a vehicle for working together on data at a distance, they could also be used to collect data from research participants in locations distant from researchers. Interaction analysis gives us leads about the effects of micro time delays and other features of e-mediated collaboration and these could be applied in testbed collaborative analysis outside present work with streaming video.

Raising the *capacity of qualitative software* to enable new standards of methodological interrelation/triangulation, and to handle data sets that exceed desktop resources, engages a number of researchers. Features to better support work with multiple data sources, including

statistical data, are a marker of current generation CAQDAS. Facilities like autocoding, table display/export, and the ability to operate other software entities such as spreadsheets in real-time within a working textual data set, already enable the use of qualitative output in multivariate techniques. AI routines prompt users to fulfill rules they have set for particular queries and estimate inferential reliability, and >merge= functions allow users to compare individual interpretations of data. But no software contains all these features, nor are there independent evaluations of them. CAQDAS for Grid use offers the opportunity to test the procedures, combine them, and apply them to larger data sets. Other capacities E-Social Science could address include exploiting non-text data sources, and escaping platform-dependency. Developments of CAQDAS are needed that better address narrative-, conversation- and discourse-analytic styles of research. A Grid application could enable CAQDAS-type work with very large textual databases such as natural language corpora. A Grid infrastructure combined with *remote sensing devices* offers new ways to understand the movement of social actors and entities such as money and commodities, capturing complex social dynamics at a time of accelerated social change. Sensor data and real-time analysis can weld information on the movement of individuals to their bank and retail data, phone and Internet logs, and health, transport and crime records. Whether this is a vision that research participants would want to see depends on the value of the resulting knowledge. Applications in social theory, planning, public health, training programmes, educational research, and participatory democracy are all foreseeable, as are applications in behavioural research, public security and the interaction of human and computational agents. But privacy concerns seem to balance knowledge gains. Exploring such dilemmas is a forté of qualitative research. Both analytic potential and ethical protections merit investigation.

Digital video streaming enables real-time interaction analysis, not only in workplace studies but in linguistic research concerned with language use in context, work on human/computer interaction (including system design), observational classroom/training studies, and visitor studies (e.g. response to museum exhibits). Findings from the CSCW field point out obstacles posed by remote co-working that make it other than a >routine practical accomplishment=, and findings from CMC research show that electronic communication has a form distinct from conventional oral and written communication. Capturing social and cognitive processes involved in real-time dispersed collaboration, and interaction in online learning environments, is seen as potentially contributing to e-Science as well as vice versa.

Digital imaging has recognised applications in healthcare and criminal justice. The digital camcorder gives qualitative researchers a means to document environmental features, and video editing and Virtual Reality software a means to manipulate the images in informative ways. Together these enable elicitation techniques that are increasingly being exploited in behavioural and social research such as in risk perception, response to the built environment, and town planning. E-Social Science also offers ways to more effectively deal with the problem of protecting research subjects= identity in work with visual data.

The Internet has prompted new *online methods for primary data collection* and E-Social Science can facilitate these. A particular impetus was provided by the interest of the product design and ICT industries in >design ethnography=. However, the innovative methods and online research tools that emerged are subject of few conventions or standards. Innovation has also proceeded in isolation and there is an opportunity to merge into an online researcher=s >tool kit= a set of applications for capturing and managing data from online settings such as Multi-User Domains and newsgroups.

As qualitative research draws on an increasing range of data sources the *visualisation of qualitative data* becomes an issue of equivalent complexity to the visualisation of quantitative data. Different representation strategies and visualisation modes need to be developed and

evaluated. *Qualitative interpretation of visual data* at a level less >micro= than that found in conversation analysis/socio-linguistics employs techniques analogous to code-and-retrieve text analysis but is ill-supported by current software. Only the simplest Boolean searches are supported, annotation for coding and memoing is limited, and documentation of techniques and technologies is lacking. Requirements for E-Social Science-based visual analysis other than for real-time distributed co-analysis need to be established.

CAQDAS may dominate current approaches to computer-supported text analysis but *XML and HTML applications* increasingly offer similar functionality and can do so in a Web environment. >Asset management software=, which can operate across a client/server-based network, is used to store and catalogue heterogenous multimedia data in such work.

Annotation by text or voice is possible, enabling coding, memoing and analytic retrievals, and Web access means that research subjects and users can participate in interpreting data. These are large advantages and, despite technical obstacles to negotiate, a promising application.

Content analysis, socio-linguistic research and work with literary and historical source material stand to gain from Grid/HPC applications that accommodate larger corpora, offer more automation, better disambiguation facilities and lexicon size, and support quasi-CAQDAS-type queries. Parallel processing could address the wrapping procedures, case processing and dynamic flexible norms that are required. There may be scope to extend extant hybrid software that amalgamates CAQDAS functionality with the dictionary-based routines of content analysis. Content analysis is not the only field that the coding paradigm does not best suit. *Systematic case-based analysis* involves quasi-mathematical qualitative methodologies for case study research where there is a need for causal explanation of outcomes. The methods and associated software are potential E-Social Science beneficiaries. A Grid research environment poses new dimensions of *ethical issues* such as informed consent, identity protection, sensitive representation and data cleaning. Data providers are increasingly cautious about Digital Rights Management, IPR and data exclusion. The activities of data intruders and the potential of remote sensing technologies challenge privacy norms. There is a need to establish baseline privacy norms and standards, explore different regulatory frameworks, and assess the impact of these and DRM on social science research. If the preceding area sounds a note of caution, that of *online publishing and dissemination* is an encouraging one to qualitative researchers. While the field is generating new forms of authorship and readership, and is concerned to reach colleagues in less-favoured research environments, especially in the developing world, e-dissemination is obstructed by lack of bandwidth and appropriate tools. Commercial publishers are also increasingly reluctant to publish monographs, data sets, or debates engendered by findings. E-publishing enables all of these. It allows data sets and analytic work to be >recoverable= by users, the linking of publications to archival material on which they are based, and the use of data mining to discover breaking topics. The technical boost offered by E-Social Science could afford the chance to test potential developments in the evolution of academic publication. Qualitative research is a useful testbed due to its exploratory orientation, and the length requirements of work with qualitative data. There are analogies for *pedagogy* in E-Social Science=s potential to promote collaboration, integrative methods, and secondary analysis.

The e-Science initiatives began with natural science, where qualitative research has done much to illuminate *the emergence and validation of scientific knowledge*, showing how technology both shapes and is shaped by social action. Qualitative analysis of the design and use of e-Science technologies could examine their effects on the conduct and content of natural science and social science knowledge. Insights from research into how qualitative researchers >do= e-mediated collaboration could inform e-Science itself, in highlighting

questions of sharing, contextuality and non-co-present interaction. The same approach could be applied to E-Social Science.

As Grid technologies, a ubiquitous digital infrastructure, quantum computing and nanotechnology move society into a pervasive computing world, there is scope for a qualitative *>futureology= study* to assess the long-term implications for social research: the potential for change, the obstacles to it, and the balance of risks and gains. Such an understanding offers social science the chance to shape the technological and regulatory framework in which research will operate. A major issue in research technology is the case for bespoke social science computing solutions and for open source software applications. Informed by the history of social science computing that comes to us largely from qualitative research, and by the distinctive strength of qualitative work as a method of discovery, qualitative research poses a singularly apt means to glimpse where our fields are going.

Recommendations for feasibility studies in qualitative E-Social Science

Our look at potential qualitative E-Social Science has intentionally been expansive, to capture what may be in frame. Inevitably there is a measure of speculation, and some potentials pose technological and/or methodological challenges that require large resources. In the dozen feasibility studies suggested below a significant part of what is needed could be addressed by modest initial support. The full report includes detailed accounts of each, and indicative *>capacity lists =* of researchers with interests in them.

Infrastructure review: the E-Social Science agenda foresees a threshold standard of IT provision, but IT for qualitative research is neither uniformly distributed nor is central computing support evenly available. The study would assess what IT resources are needed to enable participation in E-Social Science and establish what IT resources are presently available. It would also consider the scope for Grid-based resources to supplant provision of local IT resources for individual researchers.

AGNs: the study would assess the use of AGNs as a basis of research collaboration over data by distributed qualitative researchers, the potential of AGNs as a data collection facility, and the practical implications of AGN use, including installation, participation and operating costs, and booking and prioritisation issues.

A study into >scaling up=: the study would explore the scaling-up and meta-analysis of findings from cognate small-scale studies in a substantive field. It would document processes of identifying contextual features of common interest amongst different researchers, develop protocols specifying a common vocabulary for methodological and substantive facets, and assess the use of an XML meta-data model to integrate individual studies and analyses.

CAQDAS-based triangulation with Grid resources: Support for methodological interrelation is a main agenda of current generation CAQDAS but functions are dispersed between packages. The study would assess extant features and provide a requirements analysis for Grid-based/parallel processing CAQDAS able to integrate extant features and apply them to larger data volumes.

Content analysis with HPC: Natural language corpora are of interest in content analysis, linguistic research and some humanities disciplines. The study will investigate the application of HPC to such corpora by a systematic review of current automated content analysis applications and evaluate the extension of the capacities of hybrid software (offering content analysis and qualitative text analysis) by parallel processing.

HPC tools for work with digital images: Work with streaming video is only one kind of qualitative research using visual images. An important alternative involves coding and

retrieval using Boolean searches and other CAQDAS-type techniques. But existing software cannot support complex retrievals, and the granularity of coding and annotation is limited. The study would evaluate present applications and their capacities, conduct fieldwork with developers and users to specify requirements for future HPC solutions, and explore the scope of the latter to enhance digital watermarking techniques and perform more comprehensive checks against image provider databases (with a view to applications in anthropology and arts/humanities).

Online primary data collection, hypermedia and design ethnography: The Internet, and the development of >design ethnography=, have stimulated a range of online research methods and methods for using online entities as subject of research. Such work draws on several technologies to capture and manage digital data. The study would specify requirements for an online researcher= s >tool box= in relation to contemporary technologies and methods becoming available as the Internet and Grid develop.

XML for qualitative analysis: XML and associated software can incorporate qualitative data analysis functions into groupware applications and can be used across networks, a deficiency of current CAQDAS software. With preliminary development of CAQDAS functionality using XML underway, and CAQDAS developers increasingly oriented to such applications, the field is close to liberating CAQDAS from the constraints of desktop computing. The study would address access/licensing issues, optimal structure and language of XML-based applications, and technical constraints in delivering a network-based alternative to CAQDAS.

Data-sharing for secondary analysis: Qualitative secondary analysis is constrained by a culture of personal data-ownership, and a strong analytic emphasis on context. The study would assess the prospects of negotiating these constraints by peer-to-peer archiving, exploiting the analogy of a Napster-type model. A study of the implications of the Napster phenomenon for qualitative data sharing, attention to the dynamics of social reward in the open source software paradigm, and exploring the role of central data providers, could promote peer to peer data sharing as a step in increasing use of archived qualitative data.

The >massification= of qualitative research: While qualitative researchers increasingly foresee a >democratisation= of the production and output of their research, especially in multimedia-based work, we know little about whether research users or the general public are actually interested in such participation. Research in minority communities, and in online communities, does suggest research populations who seek to >speak through= qualitative research to their community and to shape research output. The study would consider what kinds of non-social scientists might participate, and in what ways, in massified consumption of qualitative research of the sort enabled by E-Social Science, and the resources E-Social Science offers to allow research users to trace how qualitative research has reached given conclusions.

Ethics and privacy in a pervasive computing world: The study would focus on aspects of a pervasive computing research environment that directly affect participation in social research (e.g., anonymity, informed consent) and document attitudes to data intrusion and to the use of sensing technologies to monitor activity. It would seek baseline data on these attitudes, norms and standards amongst different population sub-groups.

A qualitative exercise in >futurology =: E-Social Science is part of wider technological developments towards a pervasive computing environment. The study would gauge implications of such an environment for research, with a view to promoting online resources marked by social inclusion, ethical affordances, and regulatory frameworks conducive to social research. Specific issues also include the need to recursively understand the social

shaping of e-Science and E-Social Science, and the role of open source software for social science use.

Full report: Introduction and background

E-Social Science - the application of high performance computing (HPC), parallel computing and Grid technology to the research agenda of the social sciences - offers more than powerful new resources. It offers new modes of inquiry, new analytic tools, and new ways for the social sciences to engage with the societies they study. Wider developments towards a pervasive computational environment, of which E-Social Science is but a part, also pose an important object of study for the social sciences. Qualitative methods, with their openness and flexibility, and qualitative research, with its exploratory orientation, have played an important role in applying emergent technologies and in understanding the processes by which technologies are socially-shaped. UK social science has particular strength in qualitative methodology, reflected in the increasing part played by qualitative research in applied, policy-related and private sector fields as well as in academic work. Against this context it is timely to assess the potential of E-Social Science for qualitative methodology and research. This report profiles current applications of E-Social Science to qualitative research, suggests a range of potential applications, and identifies areas for feasibility studies.

Qualitative methods are used in the majority of disciplines within ESRC's remit, and in disciplines within that of the AHRB. While many researchers are committed to qualitative methodology there is a widespread view that its better integration with quantitative and IT-based methodologies would benefit substantive research fields, particularly those related to policy, such as health, criminal justice and education. The potential of >digital convergence= (Covell 1999) has been enthusiastically recognised by many qualitative researchers. The present appraisal of the potential in respect of particular methodologies and substantive areas of application will characterise current work, indicate how E-Social Science can contribute, and include details of research teams active in the area.

What is E-Social Science?

E-Social Science means different things to different researchers, but there is reasonable consensus that, at minimum, E-Social Science includes High Performance Computing (HPC), Grid computing (where numbers of computers are linked to increase computing power), and associated technologies, to support operations that desktop computing cannot efficiently provide. Generic e-Science makes a threefold distinction between the Computational Grid (primarily designed to support high performance computation), Access Grid (designed to support collaborative visualisation involving distant researchers) and Data Grid (designed for transporting large volumes of data). Among the applications these technologies offer are enhanced data sharing, enabling researchers to access remote databases via Web or Grid Services, identify studies of interest, access data sets from them and transfer these to a remote server on which analysis software is installed. Other applications include collaborative work with streaming video data, the use of Access Grid Nodes (AGNs) for dissemination and co-working, and the use of GPS-enabled handheld devices so researchers can consult remote databases while conducting fieldwork.

Some applications of Grid computing may challenge accepted data protection and ethical conventions on information collection, collation and exploitation. For example, collecting

information via remote sensors, exchanging information containing personal identifiers via personal digital devices, and compiling analyses of time-use patterns using system meta-data, brings into frame not only more finely-grained understandings of the social world but also the need for new protections for personal data security. Ethical matters are a major consideration in most qualitative methodologies and qualitative researchers have contributed significantly to present understandings of ethical research practice. Consequently an important theme relating to privacy, data collection and ethics can be contributed by *qualitative* E-Social Science to generic e-Science.

In application to humanities disciplines, e-Science has particular potential in work with archived data. Disciplines such as history have an obvious interest in enhanced means to exploit textual and other archived materials, and more recently-emergent disciplines such as dance studies have grown in tandem with developments in IT; consequently, materials such as performative corpora have from the first been held in film/video format, lending themselves both to systematic compilation as archival material and to forms of scholarship reliant on the repeated re-inspection of visual representations. Systems such as Laban notation of dancers' movements already draw on bio-engineering software that could be enhanced by greater computing power. Drama compiles film/video records of performances, and cinematic and media studies are fields where technologies relating to visual and textual archives are clearly germane. Representations of fine arts materials, including images of illuminated manuscripts and musical scores, have increasingly been made available over the Web, serving as a basis on which Grid-enabled dissemination and interpretation might build. In that E-Social Science, and e-Science more generally, challenges accepted privacy and ethical conventions, the branches of philosophy concerned with research ethics, corporate ethics, and the ethics of governance, may contribute importantly to developing regulatory and user rights frameworks for e-Science.

The present study

Research towards this report sought to establish current engagement with E-Social Science by leading qualitative researchers and to appraise potential fields of application as possible areas for feasibility studies. The report draws on response to an e-mail questionnaire, follow-up interviews, group discussions, and response by members of the AHRB e-Science committee to a preliminary document outlining qualitative E-Social Science.

An e-mail questionnaire organised into topical headings and primarily comprising free-response items was circulated in July 2003 to 60 UK and international qualitative methodologists, researchers, and computer scientists. The questionnaire indicated the ambit of E-Social Science technologies, noted central investment, invited respondents to profile current engagement with E-Social Science, and to specify planned or possible work in their field that would employ and/or assess E-Social Science techniques and technologies. It noted that ESRC may support feasibility studies in selected qualitative E-Social Science areas.

Current levels of awareness of E-Social Science argued for a sample targeted on key researchers who were active in qualitative methodology and new research techniques and

could be expected to have considered the potential of E-Social Science. Many are currently engaged in E-Social Science research. Respondents were generally experienced in applying IT in qualitative research. Easily the most-cited IT was qualitative software, with a number regularly using more than one CAQDAS package. Next most often-cited was digital video and video editing software. Digital audio was cited by several respondents, as was digital photography and image-editing software. Several cited database software, several cited text management and content analysis software, and several cited hypermedia software. There were a small number of users of mathematical modelling and simulation software, PDA-based applications, and collaborative software. Technologies each cited by one respondent included an application enabling mailing list software to be used to conduct asynchronous online focus groups, a tool supporting management and analysis of anthropological data (*anthropac*), an archiving and data description tool, audio and text digitisation software, an e-learning platform (*Blackboard*), and a tool for shared work and reader commentary in e-publishing. Nearly all respondents also taught the use of research technologies. Again qualitative software was most cited. Several taught text management and content analysis software, several the use of digital video and >visual techniques=, several XML and derivatives, and several the use of the Internet as a research medium (with one using the Internet as a language corpus). A small number taught digital audio, digital photography, hypermedia, computer-based mathematical modelling, simulation, PDA-based applications and collaborative software. A single respondent taught students about the Grid, expert systems and CATI/CAPI.

Respondents were asked to specify E-Social Science technologies that could address research problems in their field. They could illustrate responses with substantive examples and/or respond in terms of research needs in a particular substantive field such as health. It was noted that E-Social Science raises ethics and data protection issues and that respondents could respond in those terms. They could also construe E-Social Science broadly or adhere to a Grid-oriented focus. They were advised that response would be treated anonymously but, since ESRC wished to identify researchers with given interests, some might be listed; they had the option of not being listed.

The questionnaire had three sections. The >Current research technology experience= section elicited present use of research technologies and involvement in teaching their application. A section on >Current engagement with E-Social Science= asked those already engaged how long they had been involved, to describe the work, to indicate the next research challenge in progressing it, and to note any funding and whom the work involved (research students, research staff, computer scientists, non-social scientists). Respondents could answer in respect of up to three activities. The >next research challenge in this field= item gave those pursuing a substantial E-Social Science research programme the opportunity to consider its implications and what developments would be possible.

Section 3 asked respondents to identify potential contributions of E-Social Science in the next five years. They were also asked what E-Social Science technologies interested them (Access Grid Nodes; Remote access to databases; Streaming video data; Automated content analysis; Qualitative text analysis; Database capture via handheld devices; Other). A further item asked whether respondents= principal current interest in E-Social

Science was in ethics, data protection and information management and, if so, what research was needed in these areas. The section also asked whether there were research problems in their field that could not presently be addressed due to computing power constraints, a similar question being asked about technological constraints. Questionnaire response was augmented by telephone, e-mail and personal interviews with >early adopters= of E-Social Science and researchers with interests in IT applications but who were not yet engaged in E-Social Science. Group discussions involved 20 qualitative researchers from different disciplines and ranging from contract researchers to senior academics. Briefings were given on E-Social Science technologies and views sought on applications and impact on qualitative research. While those who engage with E-Social Science are generally quick to appreciate its potential contribution, no >early adopter= can be certain of the way new research technologies will make their impact felt. A measure of speculation and risk is inevitable in >blue skies= research and is a prime reason to pursue feasibility studies before committing to broader and deeper funding. It is on that basis that respondents offered their views. As one put it, >we have plenty of big questions, being tackled in a variety of ways, sometimes in piecemeal fashion. E-Social Science technologies should make it easier for these different outputs to be used, to make sharing easier and more effective=. In scope is the capacity to change the way that qualitative researchers work, the kinds of substantive problems they can address, and the way their findings can be integrated with those from other methodologies.

Existing applications of E-Social Science to qualitative methods and research

This section describes present applications of E-Social Science technologies to qualitative methodology and research. The large majority of respondents were already working on an E-Social Science activity. Most were involved in one activity, with several involved in more. Several indicated a substantial, ongoing E-Social Science research interest, particularly those involved with online databases. For example, one researcher had for several years been working on Web tools for cognitive anthropology. The project, funded by the US National Science Foundation, is building tools for Internet data collection, offering alternatives to survey questions. This project involves four anthropologists and a team of programmers.

Another database-oriented long-term activity involved a group studying linguistic morphology, examining features across genetically unrelated and geographically-dispersed languages. Most linguistic databases are >broad and shallow=, with many languages but only brief information on particular features. In contrast, this group, comprising two permanent members, three postdocs and several associates (including two computer scientists), examines relatively few languages but with depth treatment of particular features, and facilities for users to access the original data. The databases are used to create hypotheses (hence the qualitative nature of the work) and are freely available over the Web, including analytic reasoning.

Content analysis-oriented applications are not confined to comparative linguistic analysis. Anthropologists and sociologists have interests in the distribution of language forms within the same language community. Such work can have both quantitative and qualitative elements, as in a study of the interaction between social variables and the use

of swearing by different groups of English speakers, which involves analytically manipulating large natural language corpora but a qualitative style of analysis. The E-Social Science element reflects the analytic tools necessary to support the hybrid methodology and large data volumes involved. Web sites such as the DIPEX project, which collates personal experiences of major illnesses to inform patients and practitioners, offer an important research resource, forming an extensive database for narrative analysis. Software like *WinRelan* facilitates analysis of large unstructured language samples, providing automated processing, linking of knowledge-objects such as empirical generalisations, cognitive concepts, causal assumptions and indicators of value systems, output being hierarchically ordered text groups, association graphs, cause-effect structures and relevance lists (Zelger and Oberprantacher 2002). In the text analysis field, qualitative software is currently engaging with functions to support methodological interrelation (triangulation), the extension of CAQDAS to analysis of visual data, and the application of AI routines to qualitative software. An important interest is in using XML to provide CAQDAS-type tools for Web use to facilitate co-working by dispersed research teams (and research users, including research subjects).

One research team is developing technologies to support anthropological fieldwork, presently focussing on middleware. The plan is to design more effective support for the fieldwork process, including bibliographic research, interactive collection and aggregation of data, and accessing data and analysis tools at the researcher's home base from the field site. The present stage is developing XML-based Grid resources and investigating ways to reference the derivation of basic or derived data and formally relate one datum to another. The middleware is connected with context-sensitive portable devices and makes use of a >micro-Grid= of handheld PDAs. Laptops capture and produce material in the field, palmtops deploy media for interview stimulus and interactive presentations to informants. The E-Social Science challenge is to address disparate data with no fixed structure and provide ways to keep track of the application of metadata so that interpretations are traceable to original data. Obstacles remain in document management and the need for computational power when referencing documents whose high-level structures need to be re-derived in accord with analytic choices. The current project involves three researchers, two research fellows, collaboration with computer scientists, and UK cross-institutional and international collaboration.

The work links to a second activity studying the functioning of Indigenous Knowledge (IK) in bounded communities, a focus being how domain knowledge adaptively changes in response to context change, and how active use of knowledge adaptively modifies contexts of use. The technical challenge is to create specifications for tools to work with textual, visual and aural documents that operate with the middleware being developed in the first activity. There is a potential application to international development work on conserving biodiversity. The project promotes formal methods in ethnography, and, as quantitative research becomes more oriented towards agents rather than summarised parameters, this could better align these modes of enquiry. There may also be synergy between anthropologists' interests in the dynamic operation of IK systems and computer scientists operationalising expert systems that adapt to challenges from the operating environment. A related project is examining use of PDAs to exchange data from health,

social services and criminal justice databases between local agencies, so staff can track policy effects on youths at risk of developing criminal careers and identify relationships between risk and socio-economic/geographical characteristics. The research team of six includes two computer scientists, and a qualitative sociologist of technology whose role is to >improve the discourse= between social and computing scientists.

There is a significant concentration of qualitative E-Social Science in work with streaming digital video data. This builds on established interests in conversation analysis, Human-Computer Interaction, Computer Supported Cooperative Work, and workplace interaction. With applications in technologically-based industries, this work involves international collaboration, including co-analysis. It is also important in professional training relating to professional/client interaction, visitor studies in museums/galleries, and presenting >user cases= to third parties, e.g., how people use prototype software. Technological development has increased work with video data but existing software does not support an extensible, coherent, secure and distributed analytic process. It is designed for single users, not teams or dispersed co-analysts, is based on a single (often proprietary) data format, and any shared annotation facilities are >out of band= markup approaches (transcription metadata) rather than real-time annotation superimposed on the video stream itself. One group is creating Grid software to support collaborative video analysis across networks, including metadata for annotation by distributed researchers, enabling them to exploit incoming data in real time from cameras at different sites. Much of the technology is in place; the need is to specify requirements for distributed data manipulation, analysis and management, and integration of notes, transcriptions, and data from other sources. The project team comprises four social scientists, two postgraduates and four computer scientists.

While the insights of qualitative research are widely-valued, reservations about generalisability (reliability and validity) obstruct its integration with quantitative research. A key issue is extrapolation from one study to others, which requires standardised definitions and for researchers to build on each others= inquiries. Researchers studying the dynamics of family life are pursuing means to coordinate findings from small-scale studies and the computing and informatics resources (namely an XML metadata model) to support jointly rendered findings that can meet concerns over generalisability. Health research is also pursuing >meta-analysis = techniques that are more plausible to qualitative researchers than summing sample sizes from small studies and treating them as one large inquiry (an approach sometimes taken in quantitative evaluation). Protocols offering fine-grained conceptual and substantive linkage may integrate small-scale studies while increasing confidence in generalisation.

E-Social Science has potential to enhance qualitative training and networking. That is the objective of the *qualitative-research.net* online gateway, a collaboration of German universities. It offers a >link collection= of relevant Web sites (part of the SocioGuide for Qualitative Research led by the German data archive); a network mailing list; online workshop and teaching facilities; and an online journal. The initiative involves psychologists, computer scientists, humanities scholars, social scientists, and health science researchers. The *Netzwerkstatt* e-learning component supports doctoral students= online work, including an online library, mail lists and chatrooms, and facilities for

secondary analysis of online data sets. Free access to qualitative software is planned, and students can use the network to consult methodology advisors. Such environments are especially valuable where researchers work in multi-discipline/profession teams and may be isolated from other qualitative researchers. As well as universities and archives, partners include the consortium of German Academic Publishers. The latter indicates a further aspect of E-Social Science, as the online journal is to be integrated into the Social Science Virtual Library (*ViBSoz*). *ViBSoz* will provide online supplements for print publications (addressing publishers= reluctance to publish methods discussions, detailed examples and other information used in evaluating research), online tools for the publishing process, post-publication discussion media, and e-books (e.g. theses). All German social science virtual libraries will be integrated in a shared gateway in late 2003.

Providing generic resources is one way to enhance the research environment. Another is to develop resources in application to an empirical study and then engage in knowledge transfer. A project taking that approach has developed an integrated environment for data collection, analysis and authoring in a multimedia hypertext ethnography of one research site. The project is now engaged in hypertextual analysis using the authoring software *Storyspace*, along with audio/video editing software and Macromedia *Director*. The Grid element involves distributing large audio and video data sets for presentation and collaborative analysis. The team comprises six researchers, with spin-off postgraduate training. The project also has expertise in hypermedia work using *Hypercam* to record on-screen >physical= interaction within a 3D graphical environment and using mail list software (*Majordomo*) to organise asynchronous online focus groups.

E-Social Science poses access and data security issues that have been acknowledged by early adopters. For example, the sociologist evaluating communication between computing and social scientists in the at-risk youth project will examine ethical aspects of information exchange between agencies, and another project is investigating statistical disclosure risk in Grid environments, including the compatibility of confidentiality agreements across anonymised data sets (e.g. BHPS, NCDS, BCS and Census, which contain variables with different sensitivity levels and disclosure risks), and quality and attribute commonality in police, health and employee records that may be Grid networked. Grid access to multiple information sources supersedes assumptions about how much data an intruder needs to effect identification. There are large differences between states in protection for personal information, and the basis of US law in freedom of information rather than data protection, and aggressive data-seeking by multinationals, poses challenges. The project employs quantitative techniques to gauge disclosure risk in database linkage structures and will pilot methods to assess disclosure risk, but qualitative research can contribute, as noted later.

Qualitative data now covers a range of media - video, audio, text, images, graphics, fieldwork journals, databases, hyperlinks, and coding structures. Qualitative E-Social Science offers new or enhanced ways to work with such data sources. Qualitative analysis involves immersion and creative synthesis, processes that qualitative E-Social Science can aid. Surveying qualitative E-Social Science suggests that current applications include: archiving/database work; text analysis and content analysis; Grid technologies

for fieldwork; visual methods (prominently, streaming video); Access Grid Nodes as sites of co-working; teaching and publishing media; ethics/data protection/regulation.

Potential applications of E-Social Science to qualitative methods and research

This section looks expansively at possible qualitative E-Social Science applications. Asked which Grid applications currently interested them the front-runner was >Remote access to databases=, which interested nearly all questionnaire and interview respondents, followed by >Streaming video data=, with about three-quarters. About half selected >Qualitative text analysis=, >Access Grid Nodes=, >Use of Grid to enable database capture via handheld devices=, and >Automated content analysis=. Among write-ins to the >Other= category were >use of applications (e.g., simulations) that access distributed data in real time=, >distributed sensing= (remote wireless sensors) and >social/behavioural science data visualisation=. Because qualitative methods pose distinctive ethical issues it was thought that respondents might be particularly alert to ethical, data protection and information management aspects of Grid-enhanced information flows. About three-quarters had such an interest, and one felt it was reason to be sceptical about E-Social Science. Only two identified it as their primary current interest in E-Social Science.

About half the respondents thought there were problems in their branch of qualitative research that could not presently be addressed because of constraints on computing power. Among these were finding structure in unstructured data (>requiring natural language understanding, machine learning and case-based reasoning=), and the sharing, analysis, storage and distribution of large audio/visual data sets. One respondent wanted to record/replay >touch= as perceived by research subjects, but the size of data required for such sensory indicators made it impossible to record, represent and distribute across current systems. Another noted that large-scale automated content analysis exceeded desktop capacity but if researchers could manage larger data sets they could rebut criticism that studies using automated coding and analysis relied on >toy= data sets. Another noted that only limited data and queries using simplified applications were possible in real-time database work with PDAs. Re-playing these on a Grid as a server-side application would improve PDA utility in the field.

Just over half of respondents reported technological constraints other than computing power. Two saw a need to develop automated transcription systems, transcribing being a major bottleneck: >I look forward to voice recognition technology that allows us to teach a machine to transcribe text in any language (Urdu, Quechua, Hmong ...)=. Transcription software applications still require routines sufficiently labourious as to deter writing analytic memos prompted by listening to the data, and results fall short of full accuracy. Nor is there reliable information about software support for different transcription approaches. Speech recognition systems are data-intensive and much analysis needs to be done >live =; HPC could improve its capacities. Qualitative researchers studying dialogue can contribute to designing systems that interact with humans more effectively, and thus advance automated transcription software. Another identified need was for a dedicated large Beowulf cluster (parallel PC system) exclusively for social science use, and to scope the development of a version fast and compact enough to do useful work in the field.

The general view was that most agent-oriented knowledge-based applications would benefit from Grid power. >In (overseas field site) I have a telephone connection to a low bandwidth line - 64k for the whole territory (4 satellite channels)=; an individual=s share was up to 4k, rising to 16k at anti-social hours, impeding Web searching, downloading software and getting data files from collaborators. Another researcher wanted CAQDAS to provide import/export facilities >so we can get data onto the Web in a timely manner=. A humanities respondent wanted better metadata systems, including metadata content (classification schemes and ontologies) compatible with existing systems and extensible as new applications emerged. More work was needed on the interoperability of metadata systems across large, diverse collections. Also remarked were problems publishing findings from video-based research. Raw data to support analytic claims could not readily be provided, and current conventions (providing still images and transcripts) were inadequate. Nor were e-journals or providing CDs with paper publication a solution, as the materials could be downloaded and re-used for other purposes. Ways for e-journals to offer data while blocking re-use were needed.

Arts/humanities disciplines increasingly work with heterogenous multimedia data as well as traditional documentary sources, and the Arts and Humanities Data Service anticipates a five-fold increase in its current 1 terabyte holdings within two years. Practice-based arts disciplines work with records of performances or artworks, and of the creative process. Arts/humanities materials are particularly likely to have complex usage rights and/or to be commercially created/distributed, posing special access issues. The AHDS wishes to see arts and humanities virtual research archives linked to disparate resources and e-publication media, and an investigation of new forms of scholarly communication (>collaboratories=). It sees the AHDS as a testbed for such solutions, including integrating the large storage and transfer demands of the AHDS archive into the Grid. A second major interest is in establishing a metadata/ontology programme to implant a management system populated with existing thesauri on the Grid and to translate existing AHDS metadata to a Grid format to facilitate cross-domain searching. It also wishes to investigate e-publishing; ensure that common authentication/authorisation tools accommodate arts/humanities rights specifics; support exemplar projects using cross-domain knowledge, thesaurus and metadata mechanisms; and investigate automated capture of creative processes. Other humanities e-Science applications include work with digital images of manuscript sources (whose high resolution images make for large files), tools to automatically link manuscript images and transcription, and collaborative editing and annotation support. HPC could make representation of source material like handwriting and early printed books less arduous, by improved OCR systems.

We can now look at some specific potential qualitative E-Social Science applications. The potential of *applying E-Social Science technologies to archival databases* was recognised early, but secondary analysis of qualitative data raises controversial epistemological questions and remains a modest activity. Recently interest has grown in collaborative archival research and distributed analysis. This speaks to the core epistemological issue: whether qualitative researchers who did not themselves collect a given body of data can as fully understand it as would its originator. Accessing dispersed, multi-modal archival material and achieving collaboratively-produced, multiple but inter-

linked analyses would be a substantial application of Grid resources. A Data Grid would enable research teams to create and share annotations amongst themselves and external participants. Annotation requires tools distributed on the Grid ranging from simple text editors to automatic speech transcription and CAQDAS. A given data event might be represented by multiple streams, and enabling streams to be generated from annotations poses security and access control, concurrent update, and integrity rollback challenges. Annotation tools, data source format (extensibility protocols, standard format conversion), data hosting, the use of XML or relational databases to store annotations, access/security provisions, and multi-modal representation, all need assessment. While dominating discussion of qualitative secondary analysis, text is not the only archival resource. Socio-linguistics uses archived sound recordings, and ethnographic film archives offer anthropologists substantial database resources (e.g. the HADDON database). The Grid helps individuals to create and share such archives. The storage and distribution of large audio and visual data sets merits the epistemological analysis that has already been applied to text. While providing more cues than bare text, contexts of elicitation may still be hard to recover, and remote access and use pose distinctive concerns. HPC resources could address the >cherry-picking= problem in digitization projects (Banks 2001) by providing sufficient resources to make available whole collections of, e.g., ethnographic film. HPC may also help digital watermarking (e.g. *SureSign Detector*) perform more comprehensive checks against image provider databases, and permit high resolution images to be used for all work instead of only for publication (Banks 2001). HPC could also support more efficient algorithms to search visual image databases by code pattern. Despite software like *Excalibur Visual Retrieval*, most computer vision research is driven by natural science agendas and development is needed in autonomous recognition of moving images for qualitative research.

The capacity to link data is a key issue in exploiting archived data: linking qualitative and quantitative data, and linking material like personal biographies to census data, maps, and so on. There is a need to model different authentication criteria and protocols against potential patterns of use. There is a lack of suitable Web-based tools to access distributed qualitative data sources. Tools are needed that allow data to be published to the Web more easily and support online interrogation of data via standard Web browsers (equivalent to *NESSTAR Publisher*). There is limited investment in archiving suitable digital materials, and digitisation projects such as those of museums and the British Library do not focus on contemporary research materials and typically employ proprietary access systems. With these obstacles >the Grid will never be a free-for-all data sharing exercise=. Archival priorities may need re-defining for a Grid environment. *Access Grid Nodes* support meetings between remote participants and provide collaborative, computational and visualisation resources, with paralinguistic cues resulting from the large display format enabling more natural interaction. To raise awareness of these capacities AGN seminars will be presented within the ESRC Research Methods Programme by a prominent US e-Social Scientist. Testing AGNs as a collaborative mode for qualitative work could begin by applying qualitative evaluation methods to such seminars. Relevant to such evaluation is analysis of the effects of micro time-delays and the significance of visual cues noted in video-teleconferencing (Korschen et al 2002). Using AGNs as a delivery mode for seminars is a relatively

conservative application, and there is space to test their value for regular, sustained work by distributed qualitative researchers, perhaps using as vehicle an established collaboration. Since collaborative methodologies for work with streaming video are being explored in research into workplace interaction, the evaluation might involve researchers outside that field and using a different analytic approach. The scope to collect data from participants at AGN sites remote from researchers also needs evaluation.

There is a need to *evaluate the capacity of qualitative software* to support new standards of triangulation/methodological interrelation and to be applied to data sets whose size exceeds desktop computing capacities. Current generation CAQDAS offers considerable support for quantification and methodological interrelation. Command languages to automate large-scale or repetitive routines can support autocoding of participants' responses or field site information. Integrating such tools with data import/export means that quantitative data can inform interpretation before detailed coding, and qualitative output can be incorporated into correspondence analysis and multivariate techniques such as logistic regression. Such output remains linked to the qualitative data from which it originated, facilitating integration of quantitative and qualitative dimensions. Analytic retrievals can be configured using patterns suggested by quantitative analysis. Advanced tabulation features enable any number of socio-demographic features to be shown against any number of codes, with colour graduation indicating coding density (how much data supports given codes) at each cell. Some packages allow software output such as spreadsheets to be held/manipulated within coded text so users can >eyeball= output and relate it to textual data. Other packages use AI routines to prompt users to check for particular relationships, warn them they have not fulfilled >rules= they have set to support particular queries, and estimate inferential reliability. Some packages allow separately-created analyses to be merged and enable research teams to explore differences between them. Together such features offer enhanced scope for methodological interrelation, but CAQDAS for Grid use could combine them and enable their use with larger data sets.

While qualitative software has undergone much refinement, HPC could enhance CAQDAS software in other ways, such as fully exploiting non-text data sources and escaping platform-dependency. Despite promising preliminary moves, further routines are needed to facilitate collaborative analysis in dispersed teams drawing on distributed data. Narrative-, conversation- and discourse-analytic styles of qualitative research make limited use of CAQDAS, as there is little support for their notation systems nor to capture the chronological dimension; obstacles that could be overcome with HPC. Grid-based CAQDAS developments could support CAQDAS-type code-and-retrieve analysis of digital visual data. A Grid platform would enable complex retrievals interrogating very large textual databases such as the language corpora of interest to linguistic and content analysts. It would enable analysis of time-slices of online communication such as all e-mail sent on a particular date (e.g., 11 September 2001). Archival resources require an online CAQDAS-type tool to enable users to perform basic analysis before committing to download whole data sets.

Combining low-cost mobile *remote sensor devices* and an e-Science infrastructure offers new ways to track and account for the movement of actors and entities such as money

and commodities in real-time, and to develop new understandings of complex dynamics underpinning social change. Qualitative fieldwork is needed to assess the analytic returns and participants' attitudes to methods that not only involve wiring fieldsites but even issuing sensors to participants to wear. Sensor data recording the presence of individuals could potentially be related to bank data, mobile phone logs, card transactions, and transport, criminal and health records. Retailers are tagging products with Radio Frequency Identification (RFID) chips that activate when taken from shelves and send information to remote detectors, and transport providers are using them in smart cards that store records of journeys in databases. The ability to link temporal data such as call traffic and Internet usage logs could support more accurate analysis of sequential and time-based behavioural patterns. Privacy concerns and knowledge returns appear evenly balanced here, posing dilemmas of the sort best explored qualitatively. The ability to produce real-time continuous data could significantly enhance social theory, is consonant with increasing interest in performance and practice theories, and can overcome limitations of cross-sectional research. Such data enable capture of the increasingly rapid formation and evolution of communities of interest. In substantive fields such as medical interaction, classroom studies, and local democracy there is value in documenting changing practices over time and comparing participants' accounts with other records. Tracking technologies are relevant to behavioural research as well. There is scope to better understand communities that combine human and computational agents including new models of role and forms of augmented cognition. Both analytic potential and ethical protections in sensor-based observational research merit evaluation.

Qualitative research increasingly employs non-textual data sources. *Digital video streaming* enables real-time analysis of interaction, e.g., in workplace studies. For linguistic research it is also important to combine video and sound to see language use in its context, and to integrate video material with other forms of data. HCI factors in E-Social Science-based distributed data-centric collaboration pose new problems to those documented in studies of co-present desktop-based collaboration. CSCW findings show that distributed human collaboration over data is tricky. It is hard for two people in the same room to coordinate the physical manipulation of controls on a device, let alone numbers of people, dispersed, with some interested in physical manipulation, some in analytic facets, and others in the control surface. Computer-mediated interaction (CMC) mirrors neither the oral forms found in co-present interaction nor conventional written forms, and can be regarded as a post-literate form of language (or >multilogue=; Moss and Shank 2002) with distinctive features of temporality, reflexivity and community influence. Modes of communication in online learning environments can foster learning in ways not achieved in face-to-face environments. As in researching conventional classrooms, qualitative observation is method of choice in researching such environments. Research on interaction in online learning environments would also benefit evaluation of collaborative technologies such as AGNs and work with streaming video, and studies are needed of social and cognitive processes in task coordination when collaborators are not co-present. Some claims made for E-Social Science are technologically determinist and need grounding in micro-analysis of how collaboration actually proceeds.

There are similar issues in work with *digital imaging*, where a major application is to crime analysis. Experienced scene-of-crime officers see more at crime scenes than may be apparent in an image of the same. Digital imaging research is exploring whether large complex databases of arbitrary objects can be populated from multi-media-derived features (speech, transcribed language, digitised images) to enable retrieval of digitised images in response to analysts' queries. There is also work integrating digital camera and speech recognition technologies for hands-free collection and storage of images alongside collateral verbatim commentaries. Such technology would dramatically raise the detail and reliability of qualitative field observation. Digitisation also encourages re-analysis. Digital camcorders with a direct IEEE 1394 (*Firewire*) connection are attractive for fieldwork as they can download digital data direct to video editing software for conversion to a user-friendly format (e.g. *QUICKTIME*) to store in a cataloguing database. But digital video uses many gigabytes of hard disk space, so storage and computing power are obstacles. Grid solutions avoid having to compress and store on a second hard drive or burning onto CD/DVD disks. Visualisation in an HPC environment can employ virtual reality (VR) and the reconstruction of physical objects and structures. There are important applications in psychological experimentation and fields concerned with geo-spatial and built environment factors in perceived risk. Manipulating VR representations of townscapes enables elicitation from respondents of reassuring or intimidating features. HPC allows more complex structures to be presented, and more quickly. It can also overcome dangers of identifying research subjects when presenting visual data. Presently appearance is disguised by >bending= vocal pitch and/or >smudging= features - which can remove the very objects of analytic interest (e.g., eye expressions) - or by re-enactments (using actors, at some cost). The Grid makes available mass media techniques, e.g. avatars.

Both academic and applied research increasingly engage with *online methods for primary data collection*. Design and new technology industries are significant users of >design ethnography=, reflecting the surge of interest in qualitative methods in these fields. Design ethnography involves innovative visual methods and online research tools. Online research methods are still in their infancy, with few conventions, standards or methodological evaluations. E-Social Science could prompt development of an online social researcher's >tool box= merging existing technologies (like *Hypercam* and *Majordomo*) to capture and manage digital data in online settings, including Multi-User Domains, newsgroups, and other forms of online synchronous and asynchronous communication.

Data visualisation is an established challenge in quantitative research but has not been systematically explored in qualitative research. Qualitative data sets can comprise a variety of data forms, and thus be difficult to visualise comprehensively. There is a need to explore strategies for scanning, zooming, and representing textual, audio and audio-visual data, and to examine different visualisation modes (positional, temporal, spatial, thematic) to facilitate navigating and sharing data sets, which requires large storage and processing capacities.

Prevailing approaches to *qualitative interpretation of visual data* involve annotating visual images and forming sets representing different elements of an analytic classification, analogous to code-based analysis of text. Some software does allow textual

descriptions (e.g., code categories) to be appended to digitised visual images, with visual material then being recovered for analysis by similar means to the code-and-retrieve operations of CAQDAS software. However, it has been derived from other fields (animal behaviour, and marketing) and offers minimal code-and-retrieve functionality. Only the simplest Boolean searches are supported, and annotation for coding and memoing is limited. CAQDAS software like *N.Vivo* and *Atlas.ti* also enable limited hyperlinking of video segments (Secrist et al 2002). But there is no definitive account of technique/technology. There is a need to define requirements for Grid-based visual analysis software other than for real-time distributed co-analysis.

CAQDAS dominates current approaches to computer-supported qualitative data analysis but *XML and HTML applications* are increasingly able to offer similar functionality and to do so in a Web/Grid environment. Researching online accounts by cancer sufferers, Seale (2002) pre-processed the data with a customised Visual Basic program, coded using *N.Vivo* and supported analysis using a concordance generator. Such integrated software use is a pointer for the future but depends on the ability to easily import/export data. Digital convergence reinforces demand for universal, standard data formats so files can be easily transferred, and some CAQDAS packages (e.g., *Tatoo* and *Atlas.ti*) already use XML/HTML to export data files. But nearly all CAQDAS applications are customised database solutions and, because they focus on facilitating semi-automated analyses, tend not to support the range of file formats and platforms supported by mainstream database products. CAQDAS also generally lacks the client-server or three tier networked architecture to provide secure multi-user access to networked data repositories. A trait of XML is that raw data are separated from subsequent layers of markup, so source data are protected (advantageous in making analytic decisions transparent and a constraint with some CAQDAS packages). The media cataloguing database applications called >asset management software= (e.g., *Extensis Portfolio*, *iVIEWMEDIA Pro*) are well suited to storing heterogenous multimedia data. These handle a variety of file types and formats (text files, PDF, image files, audio/video), allow the original file to be opened by a chosen application for editing and analysis, and can work across a client/server based network (allowing teams to develop the same database). Documents can be annotated to create researcher-defined document sets mixing audio, text and video. Humanities researchers also need tools to enable XML markup for interpretation as well as for structural features. Multiple forms of representation help retain the >context= that is important in qualitative research. Hypermedia and multimedia data sets can include hypertext links, texts, digital audio/video, still images, annotations, graphics, diaries, coding structures and other analytical frameworks. Grid technologies are needed to handle such complex databases. Recent work on replicating the functions of qualitative software with XML is a foundation. Multimedia databases that contain data soon become unwieldy. Cataloguing databases address file size problems by storing data on a CD/DVD, other networked computers or the Web. Awareness needs to be raised of what is possible with XML and Grid resources.

Both *content analysis and socio-linguistic* research stand to gain from Grid/HPC applications. In corpus-based computational linguistics E-Social Science could accommodate larger corpora, facilitate their tagging with >off the shelf= taggers,

automate analytic operations, and improve the design of speech recognition systems, which rely on live processing and are data-intensive. Increasing amounts of textual data are available electronically, making text-based >macro studies= (large corpora, such as those amassed when analysing text-forms sampled over time) attractive in light of the value of free text over closed questionnaire items. Similar considerations apply to literary and historical source material in the humanities, where technical challenges also include material with inconsistent spelling (e.g. Early English) and non-alphabetic languages. Content analysis of natural language texts is a growth point for hybrid qualitative/quantitative methodologies. Highly-specified content analysis systems operating on macro corpora can validate small-scale qualitative work; developers of a testbed system used it to validate a qualitative study of doctor/patient interaction in cancer care. Present software to handle large corpora are limited; the requirement exceeds the capacity of CAQDAS, and most content analysis programs have limited disambiguation facilities and lexicon size (with no inbuilt dictionaries). Grid resources are needed to achieve an automated multi-functional content analysis system that can deal reliably with word ambiguity. Extensions to existing testbed software to address social science applications include case processing (to examine individual cases as well as group categories), wrapping procedures (to construct specialist dictionaries) and dynamic flexible norms (to derive standard control samples). Parallel processing procedures address such extensions. While CAQDAS organises textual data largely for code-and-retrieve analysis, and content analysis programs perform mainly quantitative dictionary-based analyses, some hybrid packages support both styles (e.g., *AQUAD*, *Textpack*). This may be a fruitful site for qualitative E-Social Science.

While the coding paradigm remains dominant in qualitative data analysis it is not best-suited to all research problems and analytic requirements, particularly causal analysis. A promising form of *systematic case-based analysis* is Qualitative Comparative Analysis, a procedure based on mathematical >truth tables = and fuzzy-set theory. Dichotomous codes are used to construct truth tables; for example, representing the presence/absence of an outcome. The truth tables represent conditions. The procedure produces one or more logically minimised sets of response configurations of causal variables associated with an outcome; systematic elaboration of outcomes against conditions enables the causally-significant conditions to be identified. QCA can be used with smaller data sets than are needed for multivariate statistical analysis, makes no assumptions about the nature of the variables, and views causation as conjectural rather than additive. It remains at core a qualitative method since field data are used to document the conditions in each case studied. The Grid could enable *QCA* software to analyse definitive volumes of cases and conditions.

Ethical issues that read differently in a Grid environment include informed consent, identity protection, sensitive representation, and data cleaning. Ethical guidelines are needed that specifically accommodate E-Social Science. Also, museums, archives and survey organisations are increasingly concerned about freely sharing information, raising issues of Digital Rights Management, IPR and data exclusion. Another concern is the economic and legal issues in combining open/shared and closed/owned systems in different partnership arrangements. New nuances of privacy, liability, risk and responsibility are raised, and their impact on public policies when everyday activities are

supported by a ubiquitous infrastructure able to monitor movement and activity. Situational crime prevention principles are already being applied to e-commerce and Internet crime, and need extension to the Grid. Means to increase the effort needed to commit crime, increase perceived risk, reduce anticipated rewards and remove excuses/justifications, merit exploration in respect of online intellectual property, intelligence, and information systems. Work on technical issues in making database structures more secure against data intruders needs to be augmented by exploratory qualitative research to establish baseline privacy norms and standards in different population sub-groups.

The development of *online publishing* suitable for presenting qualitative research is impeded by lack of bandwidth and appropriate tools. Digital technology can facilitate authorship and readership of qualitative texts, where data presentation norms increasingly diverge from publishers= commercial orientation and an expanded range of data-types and multi-layered forms of authorship, analysis and representation challenge conventional media. Grid publishing allows complex data sets and the researcher=s analytic work to be >recoverable= by users, particularly important in disciplines where providing data may be a condition of publication and in research with health implications. E-publishing sites could be linked to archives managing the data whose analysis is reported in the e-publications, and to conference papers, discussion lists and teaching resources. Data mining could establish patterns of archive use revealing emergent interests. Another issue is that Western countries train scholars from the developing world whose home institutions have poor information facilities. Grid publishing could boost the role of databases and the Internet in improving access to needed materials. Another potential application is online publication of material in minority languages. Of the 6000+ languages in the world only 279 are spoken by a million or more people. Many small languages are dying. E-publishing can help preserve such languages where commercial prospects are limited. Related initiatives include publishing in peer-reviewed print journals but posting open-access versions on OAI-compliant e-print archives for users lacking access to print media. An obstacle to take-up of the online publishing formats that can so effectively present qualitative research (with links to data, representation of additional facets by multimedia, and freer access) is the unequal status of print and electronic publications. Trialing new approaches to dissemination would help counter status problems and encourage innovation. By reducing computing resource unit costs the Grid could hasten the evolution of academic publication, and qualitative research would be a useful testbed for different models, due to the generally greater length required to present qualitative data.

Research is also needed into the implications for *pedagogy* of E-Social Science=s support for collaboration, remote data collection, use of integrative methodologies and secondary analysis. Qualitative methods have an established place in education research. Combining XML and dynamic generation of hypertext links allows the integration of qualitative data analysis techniques into more general groupware applications, making them more useful in fields like education where they can support >teacher as researcher= projects. Qualitative research has done much to advance understanding of *the emergence and validation of scientific knowledge* and e-Science is an opportunity to study such processes, showing how it evolves and what it could become. Studies of laboratory

science have made scientific norms of replicability, generalisation, and authority more transparent. Ethnographic study of the design and implementation of e-Science technologies could examine the effects of new e-technologies on the conduct and content of natural and social science knowledge. Social science has demonstrated how technology both shapes and is shaped by social action. The challenges posed by collaborative technologies in sharing qualitative data could inform the e-Science agenda itself, because they highlight questions of contextuality, ownership (of interpretation as well as data), trust, and working together that are central to the ways of working that e-Science encourages. Further, the same approach that could illuminate e-Science might be applied to E-Social Science.

As well as methodological studies, qualitative research could usefully undertake a >futuresology= study to explore the potential of Grid technologies to create a pervasive computing world, and to consider the obstacles to such a world. Qualitative methods have an accepted role in exploratory research. While enhanced speed and storage capacities are important, the essence of e-Science is the idea of pervasive computing, with access to data and applications distributed over many physical systems as if any portal was the virtual system created by the Grid. Combined with other technologies such as wireless networking, the information environment in which research is done will be pervasive in the office, home and field. E-Social Science=s enhancement of data collection and integration, layering, aggregation, collaborative analysis and dissemination will stimulate virtually integrated research centres amongst geographically isolated researchers and institutions. Quantum computing will move towards implementation over the next decade and fundamentally affect resources for social modelling and analysis. But key obstacles include agreed ethical procedures for data sharing, and restrictive Digital Rights Management agreements, in which North American and European legislation has been dominated by the interests of the entertainment industry.

Apart from >blue skies = thinking about a pervasive computing future and its effect on the research environment, there is an opportunity to shape a technological and regulatory framework that has been subject of piecemeal initiatives and contradictory trends. A prime example is that social scientists often work with proprietary and off-the-shelf technologies, for reasons of cost and funding policy. However, the history of social science computing suggests that many computational methods have arisen directly from the social science community and would be unlikely to have arisen elsewhere because they address unique requirements. Thus, while there has been much development of text-based software outside social science, with many editors, indexers, and comparison programs for program code, these are not optimised for task and social scientists have had to develop their own. *Qualrus*, a CAQDAS package significantly enhanced by AI features, was initially funded by the US National Science Foundation. *PDQ-Expert*, a tool produced by a CAQDAS developer allowing complex Census queries to be written in natural language, was funded by the US National Institutes of Child Health and Development. Some argue that research sponsors should be more involved in software development, at least reference applications. A counter-argument has been that supporting individuals to create software makes for problems of compatibility. But E-science changes the re-use value of software. Even code written for a specific specimen platform is not a barrier if the Grid includes the specimen platform among its resources.

XML, and work on application servers, can resolve many problems with transportable applications and data. The harnessing of >open source= software resolves IPR restrictions, frees users from worry that commercial software may be withdrawn, and increases support resources.

Current E-Social Science work is creating a >universal morphology= for distributed and pervasive computation but cannot specify the syntax or semantics. A >future look= could help social science influence how these technologies develop, and apply qualitative methodology=s discovery techniques to the implications of a pervasive computing environment.

Feasibility studies in qualitative E-Social Science

No new technology is wholly >transformative=, nor is its potential always immediately recognised (for example, the telephone was originally seen as a mass broadcast medium rather than a means of communication between individuals). One should not look to E-Social Science to change every area of qualitative research nor to do so overnight. However, potential applications are apparent that may benefit from feasibility studies. Among candidate areas are: an infrastructure review of IT resources for qualitative research; exploiting the potential of AGNs; a study into >scaling up=; CAQDAS-based triangulation with Grid resources; content analysis with HPC; HPC tools for work with digital images; Online primary data collection, hypermedia and design ethnography; XML for qualitative analysis; data-sharing for secondary analysis; the >massification= of qualitative research; ethics and privacy in a pervasive computing world; a qualitative exercise in futurology.

Most of the studies arise from discussion of existing/potential applications, others from background issues suggested by the applications. Different degrees of elaboration reflect progress in an area and the extent of discussion elsewhere in the report. Each study closes with names of relevant researchers. Feasibility studies might vary from a >research leave= format enabling key researchers to carry out the study, to a research contract model with the same group pursuing related studies or combining them with existing work. Feasibility studies could be linked to central facilities like the National E-Social Science Centre.

Infrastructure review: Chartered universities have pursued >device on every desk= policies to provide academics with sole-use desktop computers while in some post-1992 universities access remains poor. Central computing support for the social sciences is variable across the whole system, with site licenses seldom provided for more than two or three statistical packages and support for qualitative software ranging from the non-existent to the minimal (a situation that has changed little in a decade; Fielding and Lee 1998); other ways that qualitative researchers use technology are virtually unknown to central support services. It is important to establish exactly what IT resources are available for qualitative researchers. The more the Grid is used to make available software on networks the less important the specification of academics= desktop computers; for example, if CAQDAS-type software were available over the Grid there would be no need for researchers to have separately-licensed copies. Clearly this relies either on development of Grid-based open source CAQDAS software, a radical

improvement in XML functionality, or on developers negotiating licensing agreements for derivatives of their software to be available via access agreements to networks. Also important are institutional central computing staff with expertise in qualitative research technologies. Some institutions are exemplars in specific technologies but no institution offers comprehensive support and there are different ways this need is or is not addressed. Capacity list: Jennifer Mason (Leeds), Louise Corti (Essex), Ann Lewins (Surrey), Ray Lee (RHUL).

AGNs: The use of AGNs to support collaborative research by distributed research teams needs to be documented in a developmental demonstrator project and there needs to be an assessment of utility compared to conventional means of collaboration. While promising as a means of facilitating collaboration and collecting data as well as presenting seminar/conference talks, the distribution of AGN facilities reflects geographical accident and individual HEI policy. Despite their ability to dissolve barriers of distance, many academics wishing to use AGNs still face a substantial journey. There are also booking and prioritisation issues in using such shared resources. There needs to be a practical account of installation and running costs, although this is only an element in assessing the value of AGNs for networking between, and collecting data by, qualitative researchers. Capacity list: Keith Cole, Michael Daw (Manchester). Mike Fischer, David Zeitlyn (Kent). M. Fraser (Nottingham). B. Wessels (Sheffield). Christian Heath (KCL). Rob Witt (Surrey).

A study into >scaling up =: Sample size, non-standardised definitions, and non-cumulative patterns of inquiry currently inhibit the integration of qualitative with quantitative research, especially where the goal is triangulation (convergent validation). The feasibility study would explore the scaling up and meta-analysis of findings from cognate small-scale studies in a substantive area. XML may provide the basis of a meta-data model to integrate individual analyses. This would turn to advantage the context-sensitivity of qualitative research that is sometimes used to dismiss it as anecdotal, by identifying across sites the contextual features mutually relating to the phenomenon of interest. The approach could be used in tandem with Qualitative Comparative Analysis, a formal means to identify causal factors without complex quantification. Another approach is to build into a protocol the links between researcher, data, context and interpretation that make for context-specificity, by using an XML data model and wrappers around each individual study, so that the meta-data model can access and query individual data sets, the ownership of which remains with their originators. An ontology will be needed to specify a common vocabulary for both methodological and substantive facets; existing ontologies such as the *SRM-database* may provide a start. Capacity list: Jennifer Mason, Carol Smart (Leeds). Julia Brannen (Institute of Education). Janet Holland (South Bank University). Liz Spencer (Essex). Jane Ritchie, William O'Connor (NCSR).

CAQDAS-based triangulation with Grid resources: Grid/HPC applications could significantly enhance qualitative software by enabling it to handle larger data volumes,

promising better support for methodological interrelation. CAQDAS development has seen emergence of quasi-variable analysis pursuing regularities by using descriptive quantitative data to sort qualitative data, e.g., by exporting information about codes applied to text or a matrix built from cross-tabulated coded data. Combining such functions with a command language to automate repetitive or large-scale processes enables quantitative data to inform interpretation before detailed coding. Interrelation possibilities range from sorting qualitative comments according to categorical or scaled criteria to incorporating qualitative coding output in correspondence analysis, logistic regression or other multivariate techniques. Generating tables providing access to qualitative data from each cell of a data matrix, users can inspect any number of socio-demographic characteristics against any number of codes. Software like *Atlas/ti ver 5* enables spreadsheet (and other software output) to be >held= and manipulated in real-time within a textual project. *Qualrus* has AI routines that suggest ways to interrogate data sets, warn users if query requirements have not been met, and estimate inferential reliability. Its developer noted that *Qualrus* >will ultimately benefit from higher performance computing or Grid computing and parallel processing to examine different aspects of the data simultaneously and achieve quicker throughput from query to answer=. Several packages offer moves towards a more formal practice of qualitative analysis (and thus output more amenable to methodological integration) but these functions have not been systematically evaluated, there is no step by step account available, the packages take different approaches, and the computational requirement tests desktop computing capacity. Comparative evaluation could be used to determine how Grid-based CAQDAS could enhance methodological integration. Capacity list: Nigel Fielding, Ann Lewins, Chris Silver (Surrey). Ray Lee (RHUL). B. Wessels (Sheffield). Amanda Coffey, Beverley Holbrook, Matthew Williams (Cardiff). S. Molyneux (Sheffield). Wilma Mangabeira (Middlesex). Elaine Welsh (Oxford Brookes). Alan Bryman (Loughborough). William O=Connor (NCSR).

Content analysis with HPC: The focus is on IT support for analysis of natural language corpora using automated content analysis. There is a need to investigate the application of HPC to content analysis of macro studies of large natural language corpora and to evaluate the extension of hybrid software by parallel processing. The Communication Research Centre at Loughborough University has a comparative evaluation underway of quantitative content analysis, frame analysis, evaluative assertion analysis and discourse analysis. The Surrey Morphology Group has extensive experience of compiling socio-linguistic databases. Capacity list: Grev Corbett, Dunstan Brown (Surrey Morphology Group). Anders Hansen (Leicester). Ray Lee (RHUL). Nicholas Rossiter (Newcastle). S. Lyon (Durham). A. McEnery (Lancaster). Srikant Sarangi (Cardiff). Michael Billig, Peter Golding, Thomas Koenig, Shelley McLachlan, Katie Macmillan, Jon Potter (Loughborough).

HPC tools for work with digital images: Approaches to analysing digital visual data are not confined to those taken in researching workplace interaction. There is a different tradition based on coding and retrieval using Boolean searches. Visual images compiled using software like *Adobe Premiere* are annotated to form thematic collections (Catterall

and Maclaren 1997). Software to analyse these has been adopted from the fields of animal behaviour or marketing, and reflects these origins. For example, users train animal behaviour software (e.g. *Nodus Observer*) to recognise small numbers of non-complex behaviours that are repeated with high frequency in short time periods. Collection of behaviours is then automated and analysis is of statistical relationships showing patterns. Multimedia tools like *THEME* (Koch and Zumbach 2002) combine multivariate methods to detect behaviour patterns over time. *THEME* searches for syntactical real-time patterns based on probability theory. Applying the program to digital film, interaction patterns relating to complex behaviours were found that were not detectable by >eyeballing= the data. HPC could increase data volumes and support more valid inference from *THEME* output. With visual marketing software, digitised video data are labelled on >key= dimensions and assembled into categorical sets; analysis can be interpretive or involve tabulating categories. Complex Boolean retrievals are not supported, and the granularity of coding is restricted. An assessment is needed of computer-intensive methods for qualitative analysis of digitised visual data. As there are no published evaluations, the study would compare technical requirements and capacities, and consult actual and potential users who use different analytic approaches, to establish needs and preferences. This would identify requirements for Grid-based visual analysis software. Capacity list: Mike Fischer, David Zeitlyn (Kent). S. Lyon (Durham). S. Molyneux, B. Wessels (Sheffield). Matthew Williams (Cardiff). Marcus Banks, Sandra Dudley, Alison Petch (Oxford). Alan Cartwright (Independent consultant). Ann Lewins, Nigel Fielding, Chris Silver (Surrey). Graham Gibbs (Huddersfield). Ray Lee (RHUL). Sarah Pink (Loughborough). Miriam Catterall (Ulster). Pauline MacLaren (QUB). Helen Lomax (Luton).

Online primary data collection, hypermedia and design ethnography: New methods for collecting data online and/or using online behaviour as object of study have emerged. The application of E-Social Science to >design ethnography= has stimulated innovative visual methods and tools. These could be more fully exploited (and publicised) if gathered in an online researcher=s >toolbox= combining technologies to capture and manage digital data in online settings. Capacity list: Nicola Green, Nina Wakeford (Surrey). John Hughes (Lancaster). Claire Hewson (Bolton). Roberta Bampton, Christopher Cowton (Huddersfield). Paul Atkinson, Amanda Coffey, Bella Dicks, Matthew Williams (Cardiff). David Brown (Exeter). Mike Ball (Staffordshire). Ross Coomber (Greenwich). Nicola Illingworth (Stirling). Clare Madge, Henrietta O=Connor (Leicester). Ray Thomas (Open).

XML for qualitative analysis: XML and software to read/edit XML documents enables qualitative data analysis facilities to be incorporated into groupware applications such as collaborative workspaces and document bases and used across networks (Carmichael 2002). Collaborative systems have largely been geared to whole documents and the >pencil-level richness= of CAQDAS applications has been lacking. But when XML is combined with a scripting language like *Perl* it is possible to approach CAQDAS-type functionality (retrieval by text and codes, attachment of memos to text segments, generation of summary data) via a standard web browser. The Web delivery mode can

deepen research subjects= participation and stake in research, extending their role from providing data to participating in its interpretation. This offers potential to develop an >expert system= approach to qualitative data analysis, applying progressively-elaborated interpretive schemes to analyse new data. Data exchange and expert system capacities make XML the putative standard tool for CAQDAS-type qualitative data analysis by distributed research teams, and the Grid would facilitate the coexistence of raw data sets, proprietary CAQDAS software with XML integration and other project-specific applications. Granted publication of schemata used to structure documents, XML-based CAQDAS applications would be able to run transparently across a network. With preliminary development of CAQDAS functionality using XML underway, and a leading CAQDAS developer having set it as a distinctive orientation of his software, the field is close to liberating qualitative software from the constraints of desktop computing. Among immediate issues is the need to address access/licensing; to evaluate the optimal structure of an XML-based application; to assess languages supporting XML processing; to establish best means to code interactively, add memos to data and work with small user-defined data segments; to develop data integrity controls for co-analysis; to support complex Boolean operators; and test whether a dedicated client-side program is needed to accelerate analytic actions. Capacity list: Patrick Carmichael (Reading). David Brown (Exeter).

Data-sharing for secondary analysis: Notwithstanding the efforts of Qualidata and archives, qualitative data-sharing and secondary analysis remain constrained by a culture of personal data-ownership. A key inhibitor is the strong analytic emphasis on >context=. There is a need to appraise considerations in fitting analytic questions to data sets, and for working with incomplete data sets. Rather than rule out secondary analysis on epistemological grounds, a study into the (partial) >recovery of context= could assess what elements of qualitative data may less heavily reflect context and elicitation effects. An exploratory study could assess how sharing of materials held by individual researchers might be encouraged on a Napster-type model as part of a peer-to-peer archive. Peer-sharing is a major way that data sets come into use by other qualitative researchers, and guidelines on ownership, joint publication agreements and long-term storage may consolidate a currently informal practice onto a semi-formal level that would promote secondary analysis with the ultimate outcome of increasing use of established archives. There would be a role for a central archive to act as a >mirror= ensuring at least one copy of the data set is always available. Obstacles exposed by the Napster experience include standardisation of data formats and returns for participation. Grid mediation may address the former and the latter may be addressed by the feeling of communal endeavour involved in peer-to-peer sharing (the network of peers simply being expanded to members of a clearinghouse service). A peer-to-peer model could distribute and parallelise costs in creating central databanks to professional standard. A study of the Napster phenomenon, its application to qualitative data sharing, and attention to the dynamics of social reward in the open source software paradigm, could promote semiformal peer-to-peer data sharing as a step in increasing use of central archives in a Grid environment. Also a need is to explore IPR, ethical and commercial models of data provision in a Grid environment. A small number of providers control key data sets,

making a qualitative study, with its advantages for exploratory research, a good option to consider these and related matters (eg., ethical and authentication safeguard requirements, ownership and exploitation rights of matched data sets, and the mapping of archival resources that intersect social science and humanities interests). Capacity list: Ben Anderson, Louise Corti, Martyn Hammersley (Essex). Alan Bryman (Loughborough).

The >massification= of qualitative research: Qualitative researchers are increasingly oriented to opening their analysis and findings to research users and construing their audiences more expansively. A pluralised use of digital technology makes at least theoretically for >massification=, a >democratisation= of the production and dissemination of qualitative research. While qualitative proposals (especially multimedia-based) increasingly feature such aims and assertions we know little about whether research users (of different degrees of professional interest) or even the >general public= are actually interested in such participation. It is known that the principal readership of ethnographies, for example, is other ethnographers (Van Maanen 1988). However, there are fields where research participants are increasingly recognised as >speaking through= qualitative publications to their community, for example, among some American minority groups (Blackman 1992), it being of much interest in narrative analysis and life history method. The latest research into online >communities= (such as participants in 3D MUD virtual environments) suggests a further population interested in participating in as well as consuming the output of research (Williams 2003). For policy-makers and for other researchers it is important for qualitative research to be able to show how it has reached given conclusions, and digitally-based research enables this by providing self-documenting >audit trails= of interpretive and analytic processes and ways for research users to engage in asynchronous >dialogue= with researchers and others about the status and policy implications of particular studies or study-elements. These enablements can also >close the loop= by helping researchers to better understand the needs of research sponsors and the way research is used in the policy process. In that E-Social Science provides resources to host qualitative materials more accessibly, especially multimedia materials, we need to establish what kinds of non-social scientists might participate in massified consumption of qualitative research, and in what ways. Capacity list: Paul Atkinson, Matthew Williams (Cardiff). Martyn Hammersley (Essex). Michael Billig (Loughborough). Corinne Squire (UEL).

Ethics and privacy in a pervasive computing world: The feasibility study would explore ethical and data protection implications of altered personal data boundaries resulting from data-sharing technologies, including issues relating to research participation (e.g., anonymity, informed consent), data intrusion, Digital Rights Management, and attitudes to monitoring of activity using sensing technologies. The study would seek to establish baseline data on attitudes to ethical/DP norms and standards amongst different sub-groups of the population. Capacity list: Ray Lee (RHUL). Alan Bryman (Loughborough). Paul Thompson (Essex). Dorothy Sheridan (Sussex). Lucy Dillon (NCSR). Trudy Goodenough, Emma Williamson (Bristol). Julie Kent (UWE). Piers Benn (Imperial). Bob Brecher (Brighton). Michael Luntley (Warwick). Duncan Pritchard (Stirling). John Dixon (Plymouth). Paul Dearey (Hull).

A qualitative exercise in futurology: E-Social Science is part of wider technological developments relating not only to the Grid but the emergence of quantum computing, nanotechnology and bio-engineering, and whose effect will be a move to a >pervasive computing= environment. This will not be achieved seamlessly, due to the operation of processes of social exclusion (globally and within states), reservations about the ethics of technically-feasible innovations (e.g. surveillance use of sensing technologies), and legislative obstacles relating to DRM. Implications for the social research environment need to be anticipated and >steered=, a specific issue being support for open source software, and another being the need to understand the social shaping of e-Science and E-Social Science. Capacity list: Mike Fischer (Kent). Steve Woolgar (Oxford). Ray Lee (RHUL). Nigel Gilbert (Surrey).

References

- M. Banks (2001) Visual methods in social research, London: Sage
- M.B. Blackman 1992 >The Afterlife of Life History= Journal of Narrative and Life History, 2(1), 1-9
- P. Carmichael (2002) >Extensible markup language and qualitative data analysis=, Forum Qualitative Sozialforschung 3(2)
- M. Catterall and P. Maclaren (1997) >Focus group data and qualitative analysis programs=, Sociological Research Online, 2(1)
- A. Covell (1999) Digital convergence, Newport RI: Aegis
- N. Fielding and R.M. Lee (1998) Computer analysis and qualitative research, London: Sage
- S. Koch and J. Zumbach (2002) >The use of video analysis software in behavioural observation research=, Forum Qualitative Sozialforschung, 3(2)
- M. Korschen, J. Pohl, H. Schmitz and O. Schulte (2002) >New techniques in qualitative conversation analysis: computer-based transcription of video-conferences=, Forum Qualitative Sozialforschung, 3(2)
- C. Moss and G. Shank (2002) >Using qualitative processes in computer technology research=, Forum Qualitative Sozialforschung, 3(2)
- C. Seale (2002) >Cancer heroics: a study of news reports=, Sociology, 36(1), 107-26
- C. Secrist, I. De Kueyer, H. Bell and A. Fogel (2002) > New tools for understanding infant development in qualitative research=, Forum Qualitative Sozialforschung, 3(2)
- J. Van Maanen (1988) Tales of the Field, Chicago: University of Chicago Press
- M. Williams (2003) >Virtually criminal: deviance, harm and regulation within an online community=, unpublished PhD thesis, Cardiff: Cardiff University
- J. Zelger and A. Oberprantacher (2002) >Processing of verbal data and knowledge representation by GABEK-WinRelan=, Forum Qualitative Sozialforschung, 3(2)

Author=s note

Nigel Fielding is Professor of Sociology and co-directs the Institute of Social Research at the University of Surrey. His interests are in qualitative methods, new technologies for social research, and criminal justice. He has authored/edited books on methodological integration, qualitative software, computer technology in qualitative research, and research interviewing. He teaches qualitative methods on the UK=s longest-established social research methods programme and has been involved in the qualitative software field since the mid-1980s. He is co-editor of the *New Technologies for Social Research* book series (Sage), and reviews for *Qualitative Inquiry* (US), *Field Methods* (US), *Qualitative Research* (UK), *International Journal of Social Research Methods* (UK), and *Forum Qualitative Social Research* (Germany). He reviews in qualitative methods and new research technologies for ESRC, EPSRC, JISC, and the US National Science Foundation.