Sociological Engagements with Computing: the Advent of E-Science and Some Implications for the Qualitative Research Community

by Susan M. Hodgson and Tom Clark
University of Sheffield; University of Sheffield

Abstract

This paper explores some of the implications for qualitative researchers within sociology of developments in Grid technology and thereby aims to contribute to the debate on the future of sociological research in an increasingly digitised world. The well-established field of humanities computing provides an interesting counterpoint.

We use the methodological techniques of 'secondary analysis' and 'visual research', two currently marginal approaches within sociological research but with huge potential within e-environments, as lenses through which the potentials and pitfalls of Grid-supported qualitative work might be anticipated. Rather than a concern with the technical however, we argue for a concurrent attention to the methodological in relation to technological developments.

We find that current developments in the qualitative field are more in line with the interests of the humanities and this may shape and constrain the research that sociologists could do. Also, that the conditions to support innovative sociological developments for qualitative Grid computing are not currently well developed or supported. We conclude that in order for a more progressive e-social science agenda to emerge, a broader constituency of sociological researchers should engage with the technological debate, otherwise we risk missing out on opportunities to shape emerging technologies to our research needs.

Introduction

1.1 Discussion on the relations between computers and qualitative research practice are longstanding and can be found in both methodological and substantive literatures. Neither research practice nor computers stand still however and the need to revisit enduring concerns, as well as deliberate on novel ones, remains.

1.2 Whilst there is a growing body of research in the scientific and computing literature on Grid technology and use, the social science community has been slower to respond to these new developments. Sometimes labelled 'Internet2' or 'cyber-infrastructure', these emerging forms of distributed, computer-based infrastructures have the potential to radically alter social science research (Berman & Brady, 2005). Given how existing Internet and web-based technologies have impacted on sociological research practice, in both quantitative and qualitative domains, further discussion of upcoming possibilities seems timely. As Woolgar noted in 2003, despite potentially important developments in the way we may go on to conduct social science in the future, 'we know almost nothing about how and why (and by whom) these new technologies will be taken up, nor what will be the likely effects on the nature and conduct of e-science and e-social science research' (2003: p 2).

1.3 With this in mind, we want to explore some of the implications of the development of Grid technology by foregrounding some aspects of qualitative research practice. E-science is intended to enable and encourage particular forms of research - such as that requiring large computational capacities, collaborative work and distributed work - but is also intended to be flexible and open to the form, content and process of research. The e-science rhetoric points to the enterprise as offering a host of attractions for those who pursue knowledge regardless of data forms and we want to explore this potential critically. In this sense the paper contributes to ongoing discussions on the theme of qualitative inquiry and also the role of

computation in sociological research, including recent work utilising e-science capacities (e.g. SROa 2002 & b, Fielding & Macintyre, 2006).

1.4 We will frame our discussion in relation to existing work in humanities computing, a field with some resonance for qualitative practice. In particular, we will consider processes of 'secondary analysis' and 'visual research' as lenses through which the potentials and pitfalls of grid-supported sociological work might be anticipated. In principle, Grid technology facilitates both these processes, which are in evidence to greater or lesser extents, in existing sociological and humanistic research. The intention is not to directly compare disciplinary research practices, rather it is to use the existing experiences of humanities-driven computational work as a counter-point to our interests in these aspects of sociological practice. In using these particular examples, we hope to provoke a discussion on some of the methodological issues that will necessarily require ongoing elaboration for Grid technologies and practices, so that the technologies can evolve in ways meaningful for sociological qualitative work.

1.5 First we will provide a brief account of the Grid, e-social science and humanities computing. Next, we will consider the overlapping matters of data access, data storage and data re-use in relation to sociological and humanistic knowledge production. Secondary analysis of qualitative data and the use of visual data are two aspects of research practice on which we focus. We argue that the conditions to advance innovative sociological developments for qualitative Grid computing are not well supported and that advances in humanities work are much more developed. Following from this, we suggest that a wider appreciation of the social nature of technological development may prove an important frame for thinking about emergent technology, if the inadvertent adoption of a passive, or end-user, position is to be avoided. As such, we argue that a wider constituency of qualitative sociological researchers should engage with advanced technologies during development in order for a more progressive e-social science agenda to emerge.

The Grid

2.1 The term 'the Grid' draws from the analogy to the electric power grid. The national power grid allows you to plug in whatever appliance, wherever there is a socket, and draw electricity. Likewise, a computational grid allows you to plug in and access the processing power, resources or data that you require, wherever you, or it, are located. More specifically, Foster, Kesselman and Tuecke (2001) have described the Grid as an infrastructure that provides the means for 'flexible, secure, coordinated resource sharing among dynamic collections of individuals, institutions and resources' (p 200). A threefold distinction has emerged which differentiates between a) the Computational grid, which allows for complex and processing-hungry operations; b) the Access Grid which supports collaborative work amongst dispersed researchers; and c) the Data Grid which facilitates the moving of data between sites of storage and use. Collectively this infrastructure is envisaged to facilitate the pursuit of e-science. Each of these aspects should have distinct advantages for researchers. For example, the large increases in computational power allow for substantially improved data handling capacity; an advantage for those whose data files are numerous or large as is often the case with interviews and video data. Substantial opportunities for cross/inter/multi disciplinary work are opened up by connectivity; an increasingly visible opportunity given the encouragement of social science to work beyond disciplinary borders, the tendency towards more collaborative research and the funding opportunities offered by the EU etc. New capabilities to archive, store and access both qualitative and quantitative data are created, providing researchers today with access to previously inaccessible sources of information and potential data. As an infrastructure, the Grid is conceptually 'content-free', so the intent is for specific disciplines and fields to find ways in which to appropriate the technology to their own ends. Examples of the diversity in Grid application, across numerous academic disciplines, are easily found (e.g. RCUK, 2004).

2.2 The obvious distinction to be made here is how this is conceptually and practically different to the current usage of the Internet and the world-wide-web. This is perhaps best done by briefly exploring what is meant by the term e-science. The vision of e-science is taken to refer to science that is done through global collaborations enabled by powerful computers located at different sites that together can store and process huge amounts of all types of data (NESC, 2006). Access to larger and larger data sets and, eventually, collections of data sets, requires larger and larger computing resources. The structure of the internet, whilst useful in facilitating the exchange of relatively small pockets of information, cannot handle these larger computational requirements, primarily due to the practical limits of bandwidth and processing capability. A much larger and robust infrastructure is required to facilitate e-science. The Grid refers to the hardware, middleware, and software that makes this possible and much of this has yet to be developed. Grid technology is interesting in this regard then, as it potentially offers a host of possibilities that could tie closely to the performance of qualitative work. Perhaps even to offer a way to expand sociological engagement with computing beyond generic software, such as Word and EndNote and 'code and retrieve' packages, such as ETHNOGRAPH and NVIVO. Indeed, the Grid purports to be about more than software packages and is instead concerned with the shaping of computational architectures, within which specific

applications will emerge.

2.3 Latour (2007) has commented on how the 'click-era' of computing has created a data-rich world where our lives are increasingly being both digitised and realised. Tracing our digital footprints could produce massive amounts of interesting and productive data for sociological eyes. As he argues, the consequences for the social sciences could be enormous, in terms of data access, data volume, novel questions and our research practices. The Grid would be one way in which knowledge of our digitised world could be accomplished. But the present and the future are not the only spaces that the Grid could facilitate access to, the 'past' could also become more accessible, as we will discuss later.

E social-science and humanities computing

3.1 The ESRC participates in the overall e-science programme meaning that social science resources are already being committed, indeed national and regional centres have been established. It seems pertinent to want to understand what opportunities we have to appropriate the Grid ideals towards qualitative e-social science ends drawing on these resources. In 2003, Fielding suggested that Grid technologies could offer new modes of inquiry, new analytic tools, and new ways for the social sciences to engage with the societies they study (Fielding, 2003). A survey of prospective qualitative practice within an e-social science framework included archiving and database work, textual and visual content analysis, and, access to nodes as sites of co-working. Indeed, this latter form of working has been pursued, demonstrating that interview methods incorporating the Access Grid offered distinct, practical benefits for the conduct of work with difficult-to-reach respondents and had viable potential for international work (Fielding & Macintyre, 2006). Completion of several 'demonstrator projects' funded by the ESRC has seen the exploration, development and application of the technology within some areas of the social sciences (see NCESS, 2005 for an overview). This programme of work, and the subsequent development of further funding schemes, has been accompanied by some increase in the profile of e-social science within sociology and in the potential applications to sociologically-oriented work. However, qualitative activities are still small scale so far. A recent methods festival (ESRC, 2006) hosted five e-social science presentations, all quantitative in focus and questions remain as to whether Grid technologies will end up serving only the extension of quantitative practices, or whether the potential for qualitative researchers to shape e-science developments can be realised. Perhaps even more important, is whether Grid technologies are even a topic on the radar of the sociological community.

3.2 In contrast, humanities computing is informed by a range of qualitative practices and interests. It is a well established field with specialist journals, research centres located around the world, a variety of conferences, an identifiable workforce and other trappings that signal a disciplined pursuit. Indeed, McCarty (1999) has argued that it is a distinct discipline with over 50 years of history. Although visibly late on the e-science scene in the UK, this was due more to the late arrival of an arts and humanities research council than a lack of interest or progress in working with/on computational technologies. A wide range of projects existed prior to the advent of e-science, cutting edge methodological work is ongoing and an intriguing array of forms of research, and research practice have been enabled (see for example, http://www.oldbaileyonline.org/; http://www.hrionline.ac.uk/huntsman/index.html/; http://www.sciper.org/; http://www.ahessc.ac.uk/WorkshopsDemonstrators.html). The notion of digital, cultural heritage is strong in areas of European research (http://cordis.europa.eu/ist/digicult/index.html) and collaborative work between cultural sectors, professional organisations and academic communities has national visibility in the US (http://www.ninch.org).

3.3 Given some similarities in the pursuit of qualitative knowledge between the sociological and humanistic enterprise, there is a question as to why the humanities computing field seems more easily identifiable than a sociological computing one. To explore this issue further, we want to consider two particular aspects of qualitative work found in both research enterprises and for which Grid technologies claim huge potential, namely 'secondary analysis' and 'visual research'. Differences in the disciplinary view of these approaches allow us to explore why the differential engagement with technology may be evident. We also see how current developments within e-science and digital data-basing are already being informed by the interest and involvement of the humanities, perhaps already shaping the research we, as sociologists, could conduct.

Secondary analysis: matters of access and control

4.1 As sociological researchers, our understanding of 'secondary analysis' relates to the analysis of data which we did not collect ourselves. Often this data is to be found in existing archives, some digital, the majority not. Remarkably few methods texts pay attention to secondary analysis, the focus within sociological research being primary (empirical) data collection. Yet according to Corti and Thompson (2004, p 325) 'archived qualitative data are a rich and unique, yet too often unexploited source, of research material'. Not only can researchers pursue interests different to those sought in any original analysis,

secondary analysis can also provide additional analyses of the data or a sub-set of it as well as providing newer perspectives on any original analysis. Similarly, opportunities for extending longitudinal work, and work that attends to the temporal and/or historical aspects of social life are also areas of interest that are opened up for investigation (Hammersley, 1997). One example of the richness of such work is the Edwardians Online archive http://www.qualidata.ac.uk/edwardians/about/ that also includes some Grid-related developments http://www.esds.ac.uk/qualidata/online/data/edwardians/introduction.asp).

4.2 The Grid could open up innovative avenues to explore problems and lifeworlds that were not previously imagined to be researchable. However, such innovations are far from being unlimited and are both constrained and shaped by the social environment. Indeed, in order to discuss the constraints on its development we will now turn to the intertwined concerns of public/private data and the relationship between researcher and data. These issues around the 'conditions of acquisition' have implications for subsequent access to data whether technology is involved or not.

4.3 Usually, primary sociological data is collected and kept solely by the researcher(s) on a project. Whilst reflexive literature will frequently purport to offer an inside glimpse of particular research projects, it is not often the practice that a researcher will throw open their field notes or other records for inspection by other researchers; access to 'original' or 'raw' data is usually restricted to project participants. A researcher can be seen to own the data, in whatever form that may be. Whilst some of this data will become public in the eventual output of the study, the majority is never easily accessible or shared (see Fielding, 2000). Indeed, development work in the area of secondary analysis, for example exploring new ways of accessing and organising 'old' data, remains a marginal activity. Even QUADS, the ESRC's co-ordinating activity in this area is described as a "small initiative" in the publicity material for a showcasing event (see advert at http://www.esds.ac.uk/qualidata/news/eventdetail.asp?ID=1588).

4.4 The principles of confidentiality and anonymity are one possible explanation for the situation. The current environment of ethical practice creates conditions whereby open access to data is not entirely feasible, or even likely in the future (see Haggerty, 2004). Similarly, the intimate relationship between researcher and data could be a further reason why qualitative data are not more widely available or deposited in databanks. The notion that 'you had to be there' to be in a position to make sense of the data renders making the data available to others feel a hollow exercise (Fielding, 2000). Equally, there is also the matter of research labour, where the physical and emotional processes involved in qualitative work should not be discounted (Heaton, 1988; Hinds et al, 1997).

4.5 Whilst sociological researchers may not want to make use of earlier sociological collections, or data created by other researchers, secondary analysis could draw on any pre-existing data. Public records, image collections, and other documents such as letters, diaries or biographies, all come under the rubric of potential qualitative secondary data sources. These 'human documents', as Blumer describes them, are 'an account of individual experience which reveal the individual's actions as a human agent and as a participant in social life' (cited in Plummer, 2001, p 3). Again, however, whilst some traditions - such as symbolic interactionism - have used such data, these engagements within a more humanistic frame are often ignored by the social sciences. For Plummer, such approaches have been neglected due to a commitment to the lasting legacy of science that marginalises methods not signed up to the unity of science on one hand, and on the other, by the realist and rationalist discourse that see such documents as failing to provide any notion toward objective truth. Recent developments in post- discourse, whilst initially promising to dissolve such grand schemes, he argues, have yet to shake off such concerns (Plummer, 2001). In short, current standard research practice means that the qualitative data of sociological research is often unavailable for scrutiny by others and that other secondary sources are equally marginalised.

4.6 Research sources in the humanities, in contrast, are almost always 'open' and the origins of evidence are rarely kept hidden from view. History, for example, has always relied upon material that has been left by people in the past, who did not think of the future needs of the historian at the time they were creating it. The historian's data is largely information that has been preserved and/or re-discovered and is typically found in the public domain, in libraries, collections, museums and archives. If a source of data is held in an exclusively private domain, the work reliant on the evidence runs the risk of being dismissed as un-credible. For a historian, data (or evidence for argument) may come in the form of recorded first-hand experiences, or in the form of experiences interpreted by another person, for example a book previously written about a certain period (Marwick, 2002); there are not the same issues of ownership.

4.7 Of course many humanities researchers, and historians particularly, do not usually have the opportunity to collect data in the same and immediate way a sociologist might; their main sources of data already exist in some material form and are not constructed in the researcher-researched interaction (Katz, 2005). Indeed, the bulk of historical work will involve the analysis of data that the humanist researcher has had little control over. That said, whilst humanist research may exhibit differing ontological commitments and different research questions to the social sciences, it is noticeable that 'closeness' to the data - in the

sense of capacity to interpret it - is not a major concern.

4.8 These conditions mean that a process of continuous collation of potential research materials, and an increasing move towards digitising material, can be observed in the humanities. In contrast, whilst the major funders of sociological research, such as the ESRC, also encourage the deposition of research data within technological innovations like Qualidata[1], the majority of research within the social sciences is not funded by bodies that stipulate any guidelines on data sharing. Given that it is not a blanket requirement and does not form part of our disciplinary culture, deposition practices are likely to remain a small time activity.

4.9 In sum, a number of reasons could contribute to a culture of non-deposition of qualitative data and reticence towards secondary analysis of all types. Despite the presence of Qualidata there is often restricted availability to others' qualitative data sets. Similarly, there is a continuing lack of uptake in the use of other secondary sources. Put together with a more general reluctance toward the use of technology within qualitative research, and we find dim prospects for the development of technological tools to facilitate data sharing and secondary analyses within a sociological framework. We will point to some implications of this state shortly.

The visual in research

5.1 As with secondary analysis, the visual within sociological research occupies something of a marginal position within qualitative research practice, although it does warrant specialist journals. In 2000, Becker (p 333) pointed to the use of the visual as a signal of an 'advanced science', yet in the same year, Emmison and Smith (2000) stated that the visual remains an under-used and under-developed method of inquiry, especially when viewed in comparison to the relative success the visual has achieved in cognate disciplines. This lack of uptake within qualitative sociology exists in spite of the range of projects to which visual work can potentially contribute, as explored recently in this journal (SRO, 2005). This recent work not only focussed on the role and epistemological value of the visual, but also demonstrated the power of computational means for the representation of such research.

5.2 Indeed, there is much potential for visual inquiry in the sociological field. Emmison and Smith suggest four primary objects of inquiry for visual inquiry: firstly, the two dimensional image, which includes images, signs, and representations (cartoons etc); secondly, three dimensional data, including settings, objects and traces; thirdly, lived visual data, including the built environment and its uses; and, finally living forms of visual data, including bodies, identities and interaction. They go onto highlight how Simmel, Goffman, Elias, Foucault, and Levi-Strauss have all been 'doing the visual' when conducting research: 'taken as a whole this literature suggests that we can re-read much of the history of social thought as a tradition of visual inquiry' (Emmison and Smith, 2000: p 6). Indeed, the equally qualitative anthropological field has always relied on the visual as a source of evidence (see Bateson and Mead, 1942, for example, or Pink, 2004 for a further review).

5.3 Despite this potential, Prosser (1998) attests to the fact that there are only a small group of social science researchers interested in such research across the whole of the qualitatively minded research community and makes reference to a 'marginalisation' within the wider methods of social research. Pink (2004) also states that 'many sociologists continue to reject the use of the image in research' (p 10). When images are used within research, they are often criticised as being only cosmetic or illustrative in nature and rarely move beyond the photograph (see Becker, 1974; Ball and Smith, 1992; Emmison and Smith, 2000; and, Wagner, 2002). So, although there may be a growing interest in such research, debate about the position and status of the visual within sociology continues. Some reject it on the grounds of subjectivity, bias and specificity, some accept it providing it can be shown to be handled systematically and thus part of (a largely post-positive) social science apparatus. Others favour a more radical departure and favour the development of alternative objectives and methodologies for the use of the visual (for a fuller discussion see Pink, 2004).

5.4 In contrast, within the humanities, the oral, the written, the visual and the material can all be treated as credible 'sources' for the purpose of research: 'all these sources, although different from one another, are in many ways complementary' (Howell and Prevenier, 2001, p 27). There is no epistemological hierarchy or marginalisation of method. Using the example of history, Marwick (2002) argues that researchers must avoid taking any source as a fountain of truth. All forms of data should contend with the same methodological questions, and interpretations should be judged on their credibility (McCullach, 2004). Corroboration, not competition, is the key. Thus visual forms, whether they are photographs, cartoons, paintings, tapestries, or objects, are all equally potential useful forms of evidence, provided they are treated with similar degrees of caution and question. However, humanist researchers are not only concerned with the content and meaning of the data/source, but crucially, the context within which the data was taken and preserved. This seems the pivotal point, rather than the form the evidence takes per se. The recognition of partiality regardless of form is quite unlike sociological research, where concerns about terms such as

generalisability, validity and reliability endure and are epistemologically loaded.

5.5 The differing approaches to forms of data have an interesting and related effect upon the methodological discussions in which the different disciplines engage. Whilst qualitative sociology tends to focus upon the effect of method on the data, its subsequent analysis and ultimately its usefulness in explaining a particular area of interest, within the humanistic disciplines the back-grounding of epistemological discussions has produced an environment where methodological discussions are focussed upon finding and storing appropriate information, rather than justifying the use of specific forms. Thus key methodological concerns within the humanist enterprise centre on locating, indexing, accessibility, authenticity and credibility (Jordanova, 2000). This emphasis on finding and preserving (open) data rather than creating it has also, perhaps, meant that technology has intervened in more overt ways within in the humanist enterprise, perhaps positioning the humanities as a more powerful voice in technological debates.

5.6 This is of interest here because, as has been previously stated, there is a potential for Grid technologies to promote secondary and visual work in sociology and social sciences more generally. The development of Data Grids, for example, would assist access to multi-modal archives, regardless of location. The high performance computing power that the Grid offers could substantially increase the capacity for the storage, manipulation, sharing and analysis of visual information. However, the existence of material in digital form is a pre-requisite and digitisation projects are not an area with much sociological involvement. The implications of this have both practical and political consequences.

Digitising, databases and the will to archive

6.1 The landscape of research within the humanities had been transformed through advances in digital technology prior to the advent of Grid technologies. Over decades, humanities computing projects have built numerous editions, collections and archives of digital information. Kirschenbaum (2002) suggested that by the start of this millennium an initial plateau of first generation image-based computing projects had already been reached, with a substantial body of published and publicly accessible material already available. The massive expansion in the availability of digitised images, image collections, and other forms of representation, has also gone beyond 'official' researchers to reach others through electronic means (Jorgensen, 1999). This presence of a general readership of humanities computational work gives it added political significance in a national context of competition for resources. A project to digitise the entire collection of 18th century British parliamentary records, and give free access to all, constituted a significant enough endeavour to warrant the first UK University purchase of a fully-automated book-scanning ensemble, including a robotic page turner (Brackenbury, 2004).

6.2 Attention to a 'digitising for all' agenda means that it is not only in universities, or only for scholarly work, that computational activity occurs; museums have been in the vanguard of digitising collections for public access and educational purposes. Digital archives are increasingly becoming an important means to preserve, store and make available for re-use a wide range of data forms, including text, audio and visual media objects, for different purposes. Issues of the ongoing storage of material objects, conservation worries relating to delicate material and the increase in the educational and public remit of museums and libraries, mean that the culture of digitisation is set to grow even further.

6.3 Given this 'era of the database' (Katz, 2005, p 109) and the wealth of source material now available, often directly through the world-wide-web, the potential opportunities that exist for all researchers prepared to engage with such materials is unprecedented. Existing relations between academic humanities and the museums sector mean that digitising aims often coincide leading to provision of digital resources to serve research and education needs concurrently. In the main, however, humanities-related activity has been concerned solely with the production of databases. That is, the result of a digital project is the end collection of material and whilst some interpretation may follow, the emphasis is firmly on the provision of material. In sociological research, the production of - and access to - a database or collection is more likely to be the beginning, or only a part, of a digital project. The absence of an explicitly 'analysis' aspect to guide the design and conduct of many digital projects is an omission whose consequences are not yet known. So whilst huge effort is being expended on the detail of digital object storage and encoding (Gladney, 2004; Linden et.al., 2005), whether or not the products will be usable by sociological researchers (or others) is not currently a high priority for consideration.

6.4 This is problematic for any sociological work that aims to incorporate such databases as the mass storage of digital information creates difficulties when dealing with the existing search and retrieve paradigm that dominates social science computing. Whilst storage problems are relatively easily overcome with increasingly powerful computers, indeed the Grid is intended to facilitate such storage, the processes of locating images, data and other material is a different matter altogether. Problems of indexing and classifying research material are not new, nor the preserve of the humanities, but they are made more pressing in the digital age and take on new significance when the potential users of the material will want to

access it for different reasons and from different perspectives; as may be the case in sociological projects. As Fielding (2003) points out 'existing software cannot support complex retrievals, and the granularity of coding and annotation is limited' (p 6). Further, often 'only the simplest Boolean searches are supported, annotation for coding and memoing is limited' (ibid.). Typically, any digital representation is annotated and indexed in a word based form, it is then this that is searched rather than the image. Therefore the perceived success of any searches will depend entirely upon the would-be searcher encoding the search terms in same way as the person who encoded it in the first place. Whilst this may be less problematic in history, for example, where dates, names and provenance are often already known and can be used to initiate a search (although this is far from problematic in itself), it is likely to be hugely problematic for the social scientist who has a different way of seeing, categorising, and investigating the world. Thus we would seek to re-state Fielding's suggestion that: 'different representation strategies and visualisation modes need to be developed and evaluated' (2003, p 4). Whilst the higher processing capacity that the Grid offers could in theory support complex retrievals within and across archives, in reality these appear to be a long way off for more sociological uses. With little current engagement from the sociological community, different representational strategies that could be made possible by Grid technologies and that would better suit our needs, whatever they may be, are unlikely to be developed.

6.5 The current lack of uptake in development work so far, and the continuing marginalised position of qualitative computational methods (or use of qualitative data accessible via technology) make it unlikely that support for tentative or exploratory developments in the field will be offered much funding. Thus any developments in 'searching' technology are likely to be centred on the needs of the humanities, and ultimately the public, rather than tailored to the needs of sociological researchers. This could, and is, limiting the type of research that sociologists can conduct, and as a result will also restrict the ability of the discipline to utilise computational methods to their full extent in the future.

6.6 Huge collections of material (e.g. photo libraries) are rarely digitised in their entirety which inevitably leads to questions of what materials to digitise. Indeed, Banks (2001, p 64) presents evidence from a curator who suggests, 'while digitisation was of great value in providing access to certain collections or parts of collections, the policies behind digitisation projects often reflected a primarily content-based approach to images, a privileging of the internal narrative'. He goes on to suggest that whilst this may be helpful to some, particularly those looking for cosmetic illustrations of particular arguments, an overwhelming focus on 'pictures of…' could also lead to only certain parts of collections being digitised for reasons of cost. Such a 'cherry-picking' approach that selects 'arresting images, signature images, or images by 'names'' could directly lead to the above suggestion that certain images within collections will be digitised in preference to others. This preference for internal narrative will almost certainly privilege the storage (preservation, indexing etc.) of certain types of image over others. Collections operating under a paradigm of 'grand narratives' will select images that fit with this particular worldview. This is not only worrying for the unwitting sociological researcher who may uncritically accept resultant bias, ignorant of the selection taken place, but also for the researcher who is fully aware that images have been destroyed and can do nothing about it: 'The decisions as to what material is included in the archives, how it is classified, curated and documented, are as much social decisions as they are fiscal, pragmatic or scholarly ones' (Banks, 2001, p 104). Thus, whilst all archives and collections are necessarily constructed, how materials are indexed, selected and stored *now*, will inevitably 'condition' any future research with these materials.

6.7 Moreover, there are additional issues relating to the institutional locations, and the temporal nature, of the digitisation of archives. Fielding (2003, p 17) identified that, 'digitisation projects such as those of museums and the British Library … typically employ proprietary access systems'[2]. Whilst archives tend toward being 'open', they are frequently controlling of letting information/images go out beyond that archive. Thus, there remains a very difficult set of issues around the taking and sharing of information across a dispersed Grid framework. This is likely to remain limited without considerable involvement from the archive owners, who are not necessarily used to engaging with sociological needs and perspectives.

6.8 At a more technical level, the software needed to create even simple databases that would offer more interactive forms of analysis (beyond merely provenance information) is also not readily available on the open market nor evidently in development. This presents severe problems to any researcher looking to explore such techniques. Whilst there are developments within the digital humanities for web-crawling software that will enable the searching of multiple data-bases (see Greengrass et al, 2005, for example), computational work using archives is likely to remain difficult.

6.9 To summarise, there appear to be multiple barriers to increased sociological involvement in digital work, beyond the earlier-discussed matters of reluctance to deposit work. Pre-existing organisational relations; structural similarities in digital interests between humanities-informed communities and an emphasis on digital product, rather than analytical use of that product, all need further examination. Sociological research is less likely to be interested in the creation of digital collections than it is in the analysis of them.

However, if we are not involved in the creation, then they will not exist to be analysed. The growth in developing digital applications that relate to a more qualitative approach to understanding the world is, we believe, more likely to come from disciplines within the humanities where development and application has been established for some time and is steadily growing. As such, these humanistic interests will shape any future sociological research that utilises the subsequent digital developments.

Towards a socialised e-science?

7.1 We hope we have opened a discussion on some of the methodological issues that will necessarily require ongoing elaboration for Grid technologies and practices to evolve in ways meaningful for qualitative work. Now, we want to explore some further implications of where a continued 'lack of engagement' from sociological perspectives may lead.

7.2 There is already recognition within the humanities that computer systems need to be designed based on "principles derived from the fullest range of applicable disciplines, rather than from isolated or fragmented perspectives" (Jorgensen, 1999, p 293). Additionally:

> "HSS (humanities and social sciences) researchers should take an interest in the current Grid and e-science initiative and seek to ensure that their particular concerns as genuine potential participants are taken into account." (BA, 2005, p 63)

7.3 In this paper we have deliberately used qualitative work involving secondary analysis and use of visual data to explore whether these 'particular concerns' are indeed being addressed in a sociological context. We have suggested that there is much policy encouragement and the apparent capacity to move forwards with cross-disciplinary work in the Grid field. This seems promising; Grid developments are still in their infancy in many respects, there is much that the technology cannot do thus far, and its full potential remains unexplored - within scientific as well as social scientific arenas. It could be thought of as an 'imagined technology', with substantive aspects yet to be materialised. As one example, the drawing together of dispersed data archives to one point of access - e.g. a researcher's desk for example - presents challenging technological hurdles in the utilising of Data Grid capacities. Another is collaborative analysis of joint data sets by dispersed researchers, for example, on projects across different sites, even different countries (using computational capacity and the Access Grid). These seem to be research problems/opportunities that need to be addressed by computer and social scientists *concurrently*. Reference to existing funding programmes and ongoing projects, however, gives only limited evidence of this kind of joint activity in the qualitative domain.

7.4 Using the examples of secondary and visual methods, which have a great potential within an e-social science framework, we have demonstrated how they are both marginalised within sociological qualitative work. Within the sphere of secondary analysis, the interest in doing such analysis is limited and the deposition of material for future research access is likely to remain a small-time activity for the near future. Equally, even where an attention to the visual is explicitly paid, the emphasis remains on primary data collection rather than the re-use, initial analysis or further analysis of existing visual sources. As such, given the different orientations of various funding bodies, and the current methodological climate within sociology, both secondary research and visual work are likely to remain largely under-used and under-funded methods of qualitative inquiry.

7.5 This is in contrast to more humanist orientated disciplines that appear more encouraging toward both approaches. Here, the distinction between primary and secondary sources, and storage and analysis, is ambiguous, and there are no great distinctions to be made between the forms of data/evidence; whether they are visual, textual or oral. This results in a differing set of methodological problems and solutions with which the respective disciplinary approaches engage. One consequence is a willingness of humanistic researchers not to discount computational approaches out of hand on epistemological grounds. Given the existing prominence of technological developments within the humanities, they are perhaps more likely to drive the qualitative e-science agenda in these areas, perhaps hindering the participation of more sociological researchers and limiting the utilisation of such methods amongst the sociological community. More importantly perhaps, the products of this work are unlikely to explore the role of *analysis* within qualitative digital frameworks when the humanistic emphasis remains routed in informational search and retrieve, data management and archival access.

7.6 Fielding has already acknowledged that 'the Grid will never be a free-for-all data sharing exercise. Archival priorities may need re-defining for a Grid environment' (2003, p 17). We can speculate that, as the projects of digitisation and storage of data, including visual images, are humanities led, without considerable uptake in the use of this information amongst social scientists there is unlikely to be the funding available, nor the political will, that is needed to transform such systems.

7.7 However, this need not be the case. Although e-science will continue to be dominated and driven by science, engineering and technological research communities, Grid technologies should still remain of interest to social scientists for two key reasons. First, we can predict that the outcomes and artifacts of the e-science programme will be inextricably linked to their origins, framing and socio-political context of development (Mackenzie & Wajcman, 1999). Second, we need to contribute our expertise to the development of software and inter-faces that would best facilitate our research work, whatever its nature. Thus, the development of e-science is something we can both study and partake in.

7.8 If the social sciences, and sociology more specifically, are to influence the shaping of Grid technologies, then more work on how the Grid, and technology more widely, may be used within qualitative frameworks is needed. Furthermore, discussion of the epistemological questions raised by the interaction of the interpretive paradigm and the nomological assumptions of technological practice must continue and develop further. Indeed, there may be a need for a more identifiable social, and specifically qualitative, computing community. The advent of grid-based computational environments does open up the potential for novel qualitative engagements with computing towards the production of sociological knowledge. However, this potential should not be seen as solely a technical bandwagon to which hitching is purely optional.

7.9 If Grid technology continues to be framed as merely 'technological tools', or as 'applications' to be 'made available' to qualitative researchers, we risk being presented with a *fait accompli*. What if the resulting products don't do anything we want them too? What if they end up shaping our future research practices in ways that are uncomfortable? Akrich (1992) for example, has described how 'visions' become inscribed in technological artefacts; Adam (1998) has pointed to the inscription of gender in technical systems. We need to ask, for example, whose visions and whose gender are currently being appropriated and inscribed within the e-science system. Even computer scientists have discussed how particular sets of values become embodied in computer systems and devices (Nissenbaum, 2001). In other words, the 'social' is already, inevitably, part of e-science, but so far it is not qualitative researchers who have a voice within the e-science socio-technical system.

Final comment

8.1 Despite current engagement with e-science being a minority activity from a qualitative sociological perspective, we maintain that a widespread engagement at different levels is necessary in order to shape Grid technologies to our needs and visions. Such computational technologies should neither be shunned nor unproblematically accepted. Perhaps consideration of the practices, advantages and limitations of humanities computing is one direction in which we can look for guidance and orientation. We hinted that the humanities have the potential to dominate qualitative Grid evolution in certain areas. We believe that the failure of qualitative sociological researchers to respond to the e-science agenda risks others setting the agenda for us, both in terms of priorities and in terms of practicalities. There is a need for us to have a view on the amounts of funding, funding distributions, digitisation priorities, and overall where e-social science ought to go, for going somewhere is inevitable.

8.2 Thus we conclude that qualitative sociological researchers' engagement with e-science is necessary for political reasons as well as methodological ones. To emphasise this, we note that despite being a joint Research Councils programme, recent e-science news points to the affirmation of this as primarily a science and engineering enterprise (http://www.nesc.ac.uk/news/press_release/20060208.html). At the same time the visibility of ESRC within the overall programme remains minor, almost to the point of invisibility (http://www.rcuk.ac.uk/escience/). A serious engagement will require a wide constituency of sociological researchers to take note of these matters, not only the relatively small numbers who already work on/with technologies for research. Currently, there is a danger that, at best, the qualitative research community will experience the Grid as a passive end-user rather than as an active participant in its development. Any engagement that seeks to utilise a more innovative and progressive methodology is likely to find that the possibilities have been shaped by developments in other areas. Within sociology, engaging with e-science is a minority activity and yet widespread engagement is required in order to shape Gird technologies to our visions. Qualitative e-science is happening, and we are in danger of not being part of it until it is too late.

Acknowledgements

Notes

[1] Qualidata acts as a repository for the acquisition of data collections from qualitative and mixed methods research (for more information see http://www.esds.ac.uk/qualidata/about/introduction.asp)

[2] As well as this issue of propriety access, is the institutional focus on historical interests; 'museums and the British Library do not focus on *contemporary* research materials' (Fielding, 2003:17 emphasis added).

References

ADAM, A. (1998) *Artificial Knowing: Gender and the thinking machine* . London: Routledge.

AKRICH, M. (1992) The De-scription of technical objects. Ch.7 in Bijker & Law (ed.s) *Shaping technology/building society: studies in sociotechnical change.* Cambridge Mass.: MIT Press.

BA (2005) *E-resources for research in the humanities and social sciences* . A British Academy Review.

BALL, M., and SMITH, G.W.H. (1992) *Analyzing Visual Data.* London: Sage.

BANKS, M. (2001) *Visual methods in social research*. London: Sage.

BATESON, G., and MEAD, M. (1942) *Balinese character: A photographic analysis*. New York: New York academy of sciences.

BECKER, H. S. (1974) Photography and Sociology. *Studies in the Anthropology of Visual Communication*, 1, pp. 3-26.

BECKER, H. S. (2000) What Should Sociology Look Like in the (Near) Future? *Contemporary Sociology,* Vol. 29, No. 2, pp. 333–336.

BERMAN, F., and BRADY, H. (2005) Workshop on Cyberinfrastructure for the social and behavioral sciences: Final Report. Available from <http://director.sdsc.edu/pubs/SBE/reoprts/SBE-CISE-FINAL.pdf> [date of access 20/4/07]

BRACKENBURY, S. (2004) The 18th century British Parliamentary Papers digitisation project. A case study in the use of robotic book-scanning techniques and workflow automation to create large-scale digital libraries from bound printed texts. Available from <http://www.bopcris.ac.uk/curl_newsletter_soton_dl_2004_10_18.pdf> [date of access 4/8/06].

CORTI, L., and THOMPSON, P. (2004) 'Secondary Analysis of Archive Data' in, C. Seale et al (eds.), *Qualitative research practice*. London: Sage Publications, pp. 327-343.

EMMISON, M., and SMITH P. (2000) *Researching the Visual*. Thousand Oaks, CA: Sage .

ESRC (2006) <http://www.ccsr.ac.uk/methods/festival/> [date of access 10/11/06]

FIELDING, N.G. (2000) The Shared Fate of Two Innovations in Qualitative Methodology: The Relationship of Qualitative Software and Secondary Analysis of Archived Qualitative Data. *Forum Qualitative Sozialforschung* / Forum: Qualitative Social Research [Online Journal], Vol. 1, No. 3. Available at: <http://www.qualitative-research.net/fqs-texte/3-00/3-00fielding-e.htm> [date of access: 02/12/05]

FIELDING, N.G. (2003) *Qualitative research and E-Social Science: appraising the potential* . Manchester: National Centre for E-Social Science. Available from: <http/www.esrc.ac.uk/esrccontent/DownloadDocs/Fieldingreport.doc> [date of access: 07/12/05]

FIELDING, N.G., and MACINTYRE, M. (2006) Access Grid nodes in field research, Sociological Research Online, Vol. 11, No. 2, <http://www.socresonline.org.uk/11/2/fielding.html> [date of access 4/8/06].

FOSTER, I., KESSELMAN, C., and TUECKE, S. (2001) The Anatomy of the Grid: Enabling Scalable Virtual Organization. *International Journal of High Performance Computing Applications*, Vol. 15, No. 3, pp 200-222.

GLADNEY, H.M. (2004) 'Trustworthy 100-year Digital Objects: Durable encoding for when it's too late to ask', Available from ERPAePrints <http://eprints/erpanet.org/7/> [date of access 3/1/07].

GREENGRASS, M., HITCHCOCK, T., CIRAVEGNA, F., and CHAPMAN, S. (2005) Armadillo: Information Mining in Distributive Research Datasets in the Arts and Humanities. Sheffield: Humanities Research Institute Online Press. Available at <http://www.hrionline.ac.uk/armadillo/index.html> [date of access 07/12/05].

HAGGERTY, K.D. (2004) Ethics creep: governing social science research in the name of ethics. *Qualitative Sociology*, Vol. 27, pp. 391-414.

HAMMERSLEY, M. (1997) 'Qualitative data archiving: some reflections on its prospects and problems', *Sociology*, Vol. 31, No. 1, pp. 131-42.

HEATON, J. (1998). Secondary analysis of qualitative data. *Social Research Update*, Autumn issue. Guildford: University of Surrey Institute of Social Research.

HINDS, P., VOGEL, R. and CLARKE-STEFFEN, L. (1997). The possibilities and pitfalls of doing a secondary analysis of a qualitative data set. *Qualitative Health Research*, Vol. 7, No. 3, pp. 408-24.

HOWELL, M., and PREVENIER, W. (2001) *From reliable sources: an introduction to historical methods*. Ithaca, N.Y: Cornell University Press.

JORDANOVA, L.J. (2000) *History in Practice*. London: Arnold.

JORGENSEN, C. (1999) Access to pictorial material: A review of current research and future prospects. *Computers and the humanities*, Vol. 33, pp. 293-318.

KATZ, S. (2005) Why technology matters: the humanities in the 21st century. *Interdisciplinary science reviews*, Vol. 30, No. 2, pp. 105-118.

KIRSCHENBAUM, M.G. (2002) Image-based humanities computing. *Computing and the humanities*, Vol. 36, pp. 3-6.

LATOUR, B. (2007) 'Click-era spawns a data-rich world'. *The Times Higher Educational Supplement*, 06/04/07, p.16.

LINDEN, J., MARTIN, S., MASTERS, R., and PARKER, R. (2005) The large-scale archival storage of digital objects, Technology Watch Report, The Digital Preservation Coalition: British Library. Available at <http://www.dpconline.org/docs/dpctw04-03.pdf> [date of access 3/1/07].

MACKENZIE, D., and WAJCMAN, J. (1999) *The social shaping of technology*. 2nd Ed. Buckingham: Open University Press.

MARWICK, A. (2002) Knowledge and language: History, the humanities, the sciences. *History*, Vol. 87, pp. 3-18.

MCCARTY, W. (1999) We would know how we know what we know: Responding to the computational transformation of the humanities. Available from <http://www.cch.kcl.ac.uk/legacy/staff/wlm/essays/know/know.html> [date of access 10/11/06].

MCCULLACH, C.B. (2004) What do historians argue about? *History and Theory*, Vol. 43, pp. 18-38.

NCESS (2005), National Centre for E-Social Science (2005) *Pilot demonstrator projects* [online]. Manchester: National Centre for E Social Science. Available from <http://www.ncess.ac.uk/research/pdp/> [date of access: 25/11/05]

NESC (2006) National E-Science Centre, <http://www.nesc.ac.uk/nesc/define.html> [date of access 20/06/06]

NISSENBAUM, H. (2001) How computer systems embody values. *Computer*, 34(3), pp 118-120.

PINK, S. (2004) *Doing visual ethnography: images, media and representation in research*. London: Sage.

PLUMMER, K., 2001. *Documents of Life 2: An Invitation to a Critical Humanism.* London: Sage.

PROSSER, J. (ed) (1998) *Image-based research: A sourcebook for qualitative researchers*. Bristol, PA: Falmer.

RCUK (2004) Research Councils UK, Summary of Generic Grid Middleware and Demonstrator Projects

<http://www.rcuk.ac.uk/cmsweb/downloads/rcuk/research/esci/middleware.pdf> [date of access 20/4/07]

SROa<http://www.socresonline.org.uk/threads/qual/qual.html>

SROb <http://www.socresonline.org.uk/threads/computers/computers.html>

SRO (2005) <http://www.socresonline.org.uk/10/1/contents.html>

WAGNER, J. (2002) Contrasting images, complementary trajectories: sociology, visual sociology and visual research. *Visual studies*, Vol. 17, No. 2, pp. 160-171.

WOOLGAR, S. (2003) *Social Shaping Perspectives on e-Science and e-Social Science: the case for research support*. A consultative study for the European Social Research Council (ESRC). Manchester: National Centre for E-Social Science. Available from <http://www.ncess.ac.uk/docs/social_shaping_perspectives.pdf> [date of access: 07/12/05].