



Social Epistemology

A Journal of Knowledge, Culture and Policy

ISSN: 0269-1728 (Print) 1464-5297 (Online) Journal homepage: <https://www.tandfonline.com/loi/tsep20>

Open Access Digital Data Sharing: Principles, Policies and Practices

Natasha Susan Mauthner & Odette Parry

To cite this article: Natasha Susan Mauthner & Odette Parry (2013) Open Access Digital Data Sharing: Principles, Policies and Practices , Social Epistemology, 27:1, 47-67, DOI: [10.1080/02691728.2012.760663](https://doi.org/10.1080/02691728.2012.760663)

To link to this article: <https://doi.org/10.1080/02691728.2012.760663>



Published online: 02 Apr 2013.



Submit your article to this journal [↗](#)



Article views: 1492



View related articles [↗](#)



Citing articles: 11 View citing articles [↗](#)

Open Access Digital Data Sharing: Principles, Policies and Practices

Natasha Susan Mauthner^{a*} and Odette Parry^b

Open access to research data is increasingly being institutionalized across funding and research-related organizations through data sharing policies. Compliance with these policies has been hampered by reluctance, amongst many social and natural scientists, to release data on an open access basis. Dominant approaches to tackling data sharing obstacles focus on methodological, ethical, legal, scientific and technological infrastructures. This article explores barriers to data sharing through examination of the changing conditions and practices of data sharing; specifically, the recent introduction of digital data sharing policies and their institutionalization of open access as a normative data sharing model. The scientific, moral and political assumptions underpinning open access digital data sharing policies and principles are examined, and their implications considered for data sharing practices. The paper suggests that data sharing obstacles may be understood in terms of a policy–practice divide, whereby data sharing policies fail to recognize data sharing as a relational practice. In conclusion, alternative relational approaches to data sharing policies are proposed, drawing on the Mass Observation Archive guidelines as a useful illustrative model.

Keywords: Digital Data Sharing; Data Sharing Policies; Data Sharing Practices; Open Access to Research Data; Data Sharing Obstacles

Introduction

A major international data preservation and sharing movement is currently underway in which databases of all kinds, not necessarily created for research purposes, are increasingly being seen by governments as critical scientific resources. The availability, development and application of advanced computing and information

Natasha Mauthner is a Reader at the University of Aberdeen Business School. Correspondence to: Natasha Susan Mauthner, Business School, University of Aberdeen, Edward Wright Building, Dunbar Street, Old Aberdeen, AB24 3QY, UK. Email: n.mauthner@abdn.ac.uk

Odette Parry is Director of the Social Inclusion Research Unit at Glyndŵr University.

technology over recent years has resulted in enormous growth in the volumes of data being generated, and has facilitated and enhanced the possibilities of data preservation and sharing. The international significance of these databases has been emphasized by the Organization for Economic Cooperation and Development, a key international science policy organization, which sees them as “an essential part of the infrastructure of the global science system” (OECD 2007, 3). Nationally, the UK Government has highlighted their potential in creating “a high quality research infrastructure” (UKRDS 2009, 1; see also Economic and Social Research Council (ESRC) 2008; Office for Science and Innovation (OSI) 2007; Waller and Sharpe 2006). Data preservation and sharing is becoming a matter of national policy within and beyond the UK, including Australia, USA, Canada and Europe, where research funding agencies are responsible for developing and implementing data sharing policies (Ruusalepp 2008). Underpinning this data sharing movement is the principle of open access: the idea that “publicly funded research data should be openly available to the maximum extent possible” (Arzberger et al. 2004, 136; original emphasis).

Despite widespread institutionalization of open access data sharing policies across research funding and other agencies, uptake of these policies by researchers has been limited with reports of resistance across natural and social science disciplines (Nelson 2009). Data sharing policies, and the technologies involved in their implementation and uptake, have tended to be viewed as benign and indeed virtuous facilitators of “better” research. Resistance to data sharing has therefore been seen as a technical matter of improving data sharing infrastructures (including methodological, ethical, legal and technological elements) in order to increase compliance amongst researchers. In this paper, we challenge this conceptualization of open access data sharing policies as passive, neutral and innocent research-enhancing mechanisms. Rather, we suggest that data sharing policies, and their underlying scientific, moral and political frameworks, embody and institutionalize a particular understanding of scientific and data sharing practices that has effects. By constructing research data as public resources that are separate and independent from the human, social and material conditions of their production, data sharing policies reconfigure the conditions and practices of data sharing in ways that are in tension with researchers’ understandings of the relational and entangled nature of research data and data sharing practices.

We develop our argument in four parts. *First*, we begin with a discussion of the open access movement in an effort to situate the recent emergence and introduction of open access data sharing policies. We chart the move from open access publications to open access to research data, examine the nature of data sharing policies, and consider the case being made for open access data sharing. *Second*, we move on to discuss data sharing resistance in the natural and social sciences, highlighting similarities in the scientific, ethical and moral concerns raised across disciplinary divides. *Third*, we explore how data sharing obstacles have been tackled, and suggest that they may stem from a policy–practice divide, whereby data sharing policies fail to recognize data sharing as a relational practice. *Fourth*, we

consider **alternative** relational approaches to data sharing policies, drawing on the Mass Observation Archive guidelines as a useful illustrative model.

The Open Access Movement

Open Access to Publications

The principle of open access to publications, in its “modern” form,¹ emerged out of a grassroots open access movement originating in the physical sciences and initially aimed at increasing access to research publications. The movement started in Europe, spread to the USA where it was spurred on by the National Institutes of Health (NIH), and rapidly moved into the social sciences and humanities (Parker 2007). It was initiated in the 1960s, but became prominent in the 1990s with advances in digital technologies (Open Access Directory (OAD) 2010). A defining moment in the history of the open access movement was a conference in Budapest in December 2001, organized by the Open Society Institute, and attended by individuals representing a range of institutions with an interest in open access issues. The purpose of the meeting was “to accelerate progress in the international effort to make research articles in all academic fields freely available on the internet” (Budapest Open Access Initiative (BOAI) 2010). In 2002, the Open Society Institute launched the BOAI, a major international statement on the open access movement in which **global and free distribution of knowledge, in the form of publications, is defined as a “public good”**:

An old tradition and a new technology have converged to make possible an unprecedented public good. The old tradition is the willingness of scientists and scholars to publish the fruits of their research in scholarly journals without payment, for the sake of inquiry and knowledge. The new technology is the internet. The public good they make possible is the world-wide electronic distribution of the peer-reviewed journal literature and completely free and unrestricted access to it by all scientists, scholars, teachers, students and other curious minds. Removing access barriers to this literature will accelerate research, enrich education, share the learning of the rich with the poor and the poor with the rich, make this literature as useful as it can be, and lay the foundation for uniting humanity in a common intellectual conversation and quest for knowledge. (BOAI 2010)

The BOAI statement defines open access as free availability through the internet, and specifies what it means by “literature” in the following way:

The literature that should be freely accessible online is that which scholars give to the world without expectation of payment. Primarily, this category encompasses their peer-reviewed journal articles, but it also includes any unreviewed preprints that they might wish to put online for comment or to alert colleagues to important research findings.

The emphasis of the open access movement, as originally conceived, was focused on promoting public access to publications: documents created by researchers specifically for distribution to the public.

Open Access to Research Data²

The 1980s witnessed the application of the open access principle to research data within the natural sciences.³ GenBank was one of the earliest bioinformatics community projects on the Internet, promoting open access communications and data sharing amongst bioscientists (Benson et al. 2008). It was established in 1982 by the US NIH and is a comprehensive public database of nucleotide sequences and supporting bibliographic and biological annotation. Its establishment and near-universal use as a public or open archive have transformed the field of molecular biology and made possible the modern synthetic use of DNA sequence data. The scale of the GenBank database is a result of a communal decision to archive all DNA sequence data, a decision initially introduced by scientific journals (Whitlock et al. 2010). By the early 1990s, there was an international move towards applying the open access principle to research data within many fields of the natural sciences (e.g. National Research Council (NRC) 1997; OECD 1994; Office of Science and Technology Policy (OSTP) 1991). For example, in their study of data sharing in the natural sciences, the US NRC (1997, 10) noted that “The value of data lies in their use. Full and open access to scientific data should be adopted as the international norm for the exchange of scientific data derived from publicly funded research”. Over the past decade, open access to research data has been seen as relevant to all forms of data, disciplines and researchers across the natural sciences, social sciences, and arts and humanities. As Arzberger et al. argue, the principle of openness to research data should apply “to all science communities” (2004, 144; original emphasis) and they urge that this imperative is backed up with “formal policy frameworks and regulations” (2004, 146) by way of institutionalizing open access as a normative data sharing model.

The principle of open access is based on the notion that information is a public good which society and individuals have an obligation to make as widely accessible as possible and which individuals should be able to access as a basic right (Willinsky 2006). This principle lies at the heart of the OECD’s (2007) *Principles and Guidelines for Access to Research Data from Public Funding*, and is defined as:

the central principle that publicly funded research data should be openly available to the maximum extent possible. Availability should be restricted only by legitimate considerations of national security restrictions; protection of confidentiality and privacy; intellectual property rights; and time-limited exclusive use by principal investigators. Publicly funded research data are a public good, produced in the public interest. As such they should remain in the public realm. ... Implicitly or explicitly, this principle is recognized by many of the world’s leading scientific institutions, organizations, and agencies. Expanding the adoption of this principle to national and international stages will enable researchers, empower citizens and convey tremendous scientific, economic, and social benefits. (Arzberger et al. 2004, 136)

Open Access Data Sharing Policies

The OECD’s recommendation that publicly-funded research data be made openly accessible has led to the development of data sharing policies by research funding

agencies, scientific journals and other research-related organizations (e.g. BBSRC 2010; Medical Research Council (MRC) 2010). Research applicants are increasingly being expected to specify data storage, access and management plans, and funding agencies are instituting measures to motivate researchers to share their data such as “conditions being attached to funding schemes or ... services offered to recipients of funding” (Ruusalepp 2008, 3). Overall, funding agencies are encouraging (and in some cases requiring) their grant holders to share their data, with as few restrictions as possible, usually within a specified time frame after completion of the research (see DCC 2010; Securing a Hybrid Environment for Research Preservation and Access (SHERPA) 2009). In the UK, this is the case for the majority of the research councils, and charitable and other funding bodies are following suit (e.g. Leverhulme Trust, British Academy, Carnegie Trust, Council for British Archaeology, Wellcome Trust). Funding agency policies vary in their tone and content, and are more or less prescriptive. The UK’s ESRC, for example, requires its grant holders to deposit research data (and attaches financial penalties to non-compliance) unless there are convincing reasons for not doing so (ESRC 2010, 3–4). In contrast, the Australian Research Council (ARC) simply encourages its grant holders to consider the *benefits of data archiving*, and requests researchers provide reasons (in the study final report) where they elect not to archive (ARC 2009, 13).

The data sharing norm and ideal underlying these policies is open access: making research data *openly available to the maximum extent, and with as few restrictions as possible, through deposit in digital repositories*. In practice, variations in the *nature and sensitivity of data* are recognized and data archives maintain systems for protecting data and restricting access, where deemed necessary. In the case of *human data*, for example, these systems include: securing informed consent for data preservation and reuse from participants at the time of data collection; removing personal identifiers and sensitive information through anonymization procedures; and implementing legally-binding license agreements specifying the rights and responsibilities of depositors, archives and end users, particularly with respect to ensuring confidentiality of the data. Depositors can also impose specific conditions of access to, and use of, the data (see Bishop 2009).

The *time frames* for data sharing vary across policies. They range from three months of the end of an award (e.g. ESRC) to 6 months (e.g. ARC, European Research Council (ERC)), 12 months (e.g. Research Foundation Flanders) and three years after project completion (e.g. Biotechnology and Biological Sciences Research Council). In some cases, the time frames specified are less restricting and/or more open to interpretation. For example, some require deposit “within a reasonable time after completion of the work” (e.g. NERC, MRC, Health Education Authority), “immediately after publication of results” (e.g. Cancer Research UK, Canadian Institutes of Health Research (CIHR), Wellcome Trust, Genome Canada) or “at the earliest possible opportunity” (e.g. Gordon and Betty Moore Foundation, USA) (see SHERPA 2009 for further details).

A further trend is the introduction of data archiving policies by scientific journals, making publication conditional on data archiving. This is a well-established tradition in fields such as crystallography (Arzberger et al. 2004) and molecular biology (Benson et al. 2008), and has recently been introduced by several key journals in the fields of evolution and ecology (Bruna 2010; Whitlock et al. 2010).

The Case for Open Access Data Sharing

The case for open access to research data is made on the basis that it delivers scientific, moral, economic, political, professional, social and security benefits (see Arzberger et al. 2004; Canadian Association of University Libraries (CARL) 2009; OECD 2007; Social Sciences and Humanities Research Council (SSHRC) 2009; UKRDS 2009).³ Re-use of digitally-preserved data is regarded as an efficient and cost-effective use of public funds. It is seen as promoting the global empowerment of researchers and scientific communities, and the democratization of knowledge. Data depositories are seen to provide an important resource for training in research. Data reuse is understood to reduce the burden on respondents and communities caused by multiple data collection efforts. Data storage is considered to reduce the information security risks associated with maintaining duplicated data-sets in more than one location.

Scientifically, open access data sharing is understood to promote transparency, innovation and progress (Arzberger et al. 2004). It is seen to support open scientific inquiry and improve methods of data collection through the scrutiny of others (Bruna 2010). It is perceived as providing a means for scientific innovation by promoting new research, new questions, new methods and new insights. Topics not envisioned by the initial investigators can, it is suggested, be explored and new areas can be identified for collaborative study. The ability to retrieve and compare data from multiple sources can lead to the testing of new or alternative methods, and to “powerful new insights” (UKRDS 2009, 1). Combining data-sets from different departments, agencies and sources is seen to allow the creation of new data-sets which can facilitate high-quality, policy-relevant research by providing a fuller picture rather than analysing separate pieces of a jigsaw (ESRC 2008).

The moral case for open access derives from arguments that publicly-funded researchers have a moral duty to the public to make available data collected using public funds (Bishop 2009). As the MRC (2010) policy on data sharing explains “publicly-funded research data are a public good, produced in the public interest, and ... they should be openly available to the maximum extent possible”. And Arzberger et al. (2004, 135) comment that:

In recent years, the debate on e-science has tended to focus on the “open access” to the digital output of scientific research, namely, the results of research published by researchers as the articles in the scientific journals. ... This focus on publications often overshadows the issues of access to the input of research—the research data, the raw material at the heart of the scientific process and the object of significant annual public investments.

Research data are therefore being defined as public resources and public property. The UK Research Data Service similarly positions research data as research outputs: “research data will increasingly be the starting point for new research as well as a key output” (UKRDS 2009, 1; see also CARL 2009, 4). Moreover, as moral agents, researchers are cast as custodians of public property whose role is one of stewardship of public knowledge (CARL 2009; MRC 2010; RIN 2010). This stewardship serves to ensure that “research data are easily accessible, so that they can be used as often and as widely as possible” (Arzberger et al. 2004, 135).

Resistance to Open Access Digital Data Sharing

In practice, open access data sharing policies have had limited impact across disciplines and nations (CARL 2009; Heaton 2004; RIN 2009). Studies report refusal, avoidance and delay vis-à-vis data archiving/sharing amongst researchers; and the frequent use of researcher-imposed restrictions whereby open access is granted to data summaries but prohibited to raw data-sets (e.g. Arzberger et al. 2004; Bishop 2009; Cambon-Thomsen, Rial-Sebbag, and Knoppers 2007; Hine 2006; Nelson 2009; Zimmerman 2007). Examples of data sharing obstacles from the natural and social sciences highlight similarities in the scientific, ethical and moral concerns raised across disciplinary divides.

Examples from the Medical and Natural Sciences

Genomics is an example of a field of study where ambivalence about data sharing is evident. Genomics has enjoyed some spectacular successes as a result of open access data depositories such as GenBank, one of the earliest bioinformatics community projects on the Internet, which promotes open access communications and data sharing amongst bioscientists (Benson et al. 2008). GenBank comprises a comprehensive public database of nucleotide sequences and supporting bibliographic and biological annotation, and its near-universal use as a public archive has transformed the field of molecular biology and made possible the modern synthetic use of DNA sequence data. The scale of the GenBank database is a result of a communal decision to archive all DNA sequence data, a decision initially introduced by scientific journals (Whitlock et al. 2010).

Although Genbank has been used as a paradigmatic model for data sharing policies and practices (RIN 2009), recently questions have been raised about the archiving, sharing, reuse and control of genotypic and clinical data. One concern is how data can be shared in the interests of science and society, and also protect the interests of researchers and their respondents (Kaye et al. 2009; O'Brien 2009). The concepts of informed consent and anonymization have to date provided a normative ethical framework for addressing issues of respondent confidentiality, privacy and protection. However, these concepts are increasingly being problematized. Informed consent has been criticized as inadequate for addressing ethical concerns because in practice it is satisfied by “largely ignorant blanket permission”

(Greely 2007, 361). Achieving the level of understanding that is required for truly informed consent in a data sharing context is difficult. Researchers can provide only limited information and reassurances about potential future uses and users of shared data (Greely 2007). Placing data within a public digital archive can mean that both researchers and respondents lose control over usage of the data. Indeed, retrospective discovery by research participants of unknown or unconsented applications of their data has led to legal disputes (O'Brien 2009). A further issue concerns national variations in ethical norms, guidelines and practices which may render data reuse across national borders unethical and illegal. For example, the western tradition of informed consent operates according to the principle that the most specific consent is the best consent. When data are exchanged across national boundaries, there may be little scope for regulating whether data reuse conforms to the specific purposes that respondents have consented to (Cambon-Thomsen, Rial-Sebbag, and Knoppers 2007).

While conventional wisdom is that **anonymization of participant information**, and the removal of obvious personal identifiers, provides effective privacy protection, recent evidence suggests this is not necessarily the case. Genome-wide scans provide such a wide range of unique data points that an individual can be identified in aggregate data (Kaye et al. 2009; O'Brien 2009). Because "absolute anonymity is basically unachievable in these large cohort population databases" (O'Brien 2009, 201), the whole concept of anonymization is under question (Kaye et al. 2009; O'Brien 2009). Indeed, there is a call for cohorts and biobanks to abandon anonymity for biospecimens. Greely (2007), for example, argues that anonymization is undesirable and unethical because it fails to adequately protect subjects. It effectively takes away participants' rights to withdraw from a study, makes it harder for researchers to re-contact donors, and makes it easier for researchers and Biobank managers to overlook the problems of protecting donor information. In light of these critiques, some researchers are calling for a more cautious approach to public data release, suggesting that "investigators and research ethics boards need to carefully consider whether public data release is warranted" (Cauldfield et al. 2008, 434). They argue that restricted, rather than open, access may be a more appropriate normative model (Cauldfield et al. 2008).⁴ Indeed, some funding agencies are now stalling Biobank projects, while others have withdrawn genomic data from public archives. In 2008, the US NIH terminated all open access to the NIH Genome-Wide Association Study database and recommended that other institutions do the same. The Wellcome Trust has taken similar steps with the UK Biobank Web access (Couzin 2008; O'Brien 2009).

Several genomics researchers have also raised scientific concerns over biobank projects. They argue that procedures employed to protect the identity of each data subject may lessen the scientific value and utility of the data (Greely 2007; Kaye et al. 2009). Some also object to the model of science underpinning the establishment of genomic databanks, and its assumed universal scientific validity. They argue that biobanks embody a reductionist and geneticized approach to health and epidemiology that is controversial and contested (Wallace 2005).

The rights and interests of the researchers who generate the data have also become a focus of attention (e.g. Hine 2006; Kaye et al. 2009; O'Brien 2009). For researchers, the collection of data entails significant personal and professional investments, and there appears to be “a career imperative at the heart of this resistance to open sharing of data” (RIN 2009, 38–9; see also Bruna 2010; Zimmerman 2007). Some genomics researchers are concerned about how data generation, a skilled practice that requires time and career investments, will be recognized and rewarded in its own right. They argue that researchers have intellectual property rights—moral ownership, priority, use and control rights over their data—that are inadequately recognized within open access policies.

These concerns about open access are echoed more broadly across the life sciences (from botany, to ecology, to clinical neuroscience), where researchers have expressed unease about potential misuse of data, ethical constraints and intellectual property (RIN 2009). Data collection can be difficult, expensive and time-consuming, particularly when it involves building relationships of trust with other stakeholders or when it requires the development of innovative techniques. Researchers understand data as part of their intellectual capital and as critical to securing career advancement, rewards and recognition. They are “wary of giving away their data for someone else to analyse and get the credit” (RIN 2009, 38). Researchers have expressed a sense of ownership (e.g. Zimmerman 2007), protectiveness and responsibility towards their data, associating open access with a loss of control over their data and how they might be re-used.

Finally, physicists working at the European Organization for Nuclear Research (CERN) have signalled reluctance to place “raw” data in public archives. They argue that these data have limited scientific value when taken out of their experimental context. Moreover, science is a competitive enterprise and “raw” data provide the intellectual capital scientists require to compete. While CERN overtly subscribes to the principle of open access, in practice, “raw” data are protected and only made available to other researchers on a negotiated basis. Only “cleaned up and interpreted data” are openly accessible to the international physics community (Arzberger et al. 2004, 145).

Examples from the Social Sciences

In the social sciences, similar ambivalence has been expressed by qualitative researchers within and beyond the UK about emerging data sharing policies (e.g. Broom, Cheshire, and Emmison 2009; Perry 2008). In the UK, this ambivalence is evidenced by requests for exemption from archiving by researchers funded by the ESRC. The most frequent reasons given are ethics-related (e.g. lack of consent, confidentiality), which affect up to 25% of qualitative collections offered (Bishop 2009; see also Perry 2008). These concerns indicate striking parallels with those of natural scientists, in questioning the ethics of informed consent (Richardson and Godfrey 2003), particularly within the context of international research where informed consent and data archiving and sharing may have culturally-specific meanings (Morrow 2009).

Similar to colleagues in the natural sciences, some social scientists have problematized the assumption that anonymized data necessarily retain their scientific integrity, and question the epistemic integrity of using archived data out of their contexts of production (Broom et al. 2009; Hammersley 1997). They too highlight researchers' moral and legal ownership of research data, their intellectual property rights and their moral rights to priority use of the data. Given that qualitative research involves the disclosure of personal information on the part of researchers, some social scientists raise concerns over how researchers' personal and professional rights and interests will be safeguarded in an open access context (McLeod and Thomson 2009). They draw additional attention to the power differentials between potential data "users" and the data "suppliers" (Griffin 1997); the potential impact of data sharing policies on research practices (e.g. willingness of respondents to participate in research; nature of researcher and respondent narratives); and the erosion of academic freedom and autonomy through infringements on researchers' rights to determine appropriate data sharing activities and practices.

Open Access and the Reconfiguration of Data Sharing Contexts and Practices

Resistance to data sharing policies amongst academic communities has prompted much debate about the obstacles and how to overcome them. Attempts to address barriers to open data sharing have tended to view them in computational or operational terms: as amenable to technical fixing. The focus has therefore been on improving methodological, ethical, legal, scientific and technological infrastructures. For example, ethical issues have been addressed through the development of more sophisticated ethical protocols for data reuse which are understood to ensure better compliance with ethical guidelines (regarding informed consent and data anonymization) and legal requirements for data sharing, and restriction of data reuse to "bona fide" researchers (e.g. Bishop 2005, 2009; Cambon-Thomsen, Rial-Sebbag, and Knoppers 2007). Similarly, the formulation of data standards and the requirement to archive contextual information (e.g. through metadata, hypertext or hypermedia) have been seen as important in rendering "raw" data more meaningful by increasing their representational accuracy (e.g. Bishop 2006; Dicks et al. 2006; Moore 2007; Gardner et al. 2003; van den Berg 2005).

These attempts to overcome data sharing obstacles have not had the desired impact, and policy-makers, research agencies, data archivists and scholars advocating open access continue to express concerns over limited compliance amongst academic communities. Resistance to data sharing policies is all the more intriguing given that there is generally widespread support across disciplines for the principle of data preservation, sharing and reuse and its scientific potential (e.g. Broom et al. 2009; Gardner et al. 2003; Perry 2008; RIN 2009). The natural sciences have long enjoyed a strong culture, ethic and tradition of data sharing (RIN 2009). Currently, many science researchers, such as those from the historical and comparative sciences (including geology, paleontology, museology, taxonomy,

archaeology, anthropology, history and the archival sciences), make routine use of data collected by others. Likewise, in the social sciences, and arts and humanities, establishment and use of archives, preservation of data in the form of personal papers, government records and historical and cultural artefacts, and secondary use of quantitative data-sets are all well-established practices (Valge and Kibal 2007). Indeed, it is recognized by researchers across disciplines that data management repositories, or similar data preservation support services, are key to the preservation of research data beyond the life of specific projects and publication cycles.

One way of furthering our understanding of resistance to data sharing policies is by focusing on the policies themselves and how they are changing data sharing conditions and practices. For example, whereas data sharing was once a matter of individual choice, the introduction of data sharing policies means that researchers are increasingly expected, and in some cases required, by higher authorities to share their data. Furthermore, data sharing policies promote, and in some cases prescribe, a specific data sharing norm: that of open access. Finally, whereas data sharing has traditionally taken place through personal exchange, online databases mean that digital data can be instantly shared with the wider public and scientific communities (subject to access conditions). The introduction of data sharing policies is therefore reconfiguring scientific practice and research relationships, including those between and amongst researchers, respondents, research funding agencies, research institutions and the public. Specifically, these changes entail an erosion of researchers' control over data sharing conditions and practices, and less discretion to decide whether, when, how, what and with whom to share their data. In order to understand these processes further, it is useful to explore how open access data sharing policies embody and institutionalize particular scientific, moral and political frameworks (and ideologies).

The Scientific Framework Underlying Open Access Data Sharing Policies

The scientific case for open access data sharing is underpinned by a particular model of science that can be gleaned from the way in which "data" are defined and understood within two foundational documents on data sharing. These are Arzberger et al.'s (2004) paper on "Promoting access to public research data for scientific, economic, and social development" and the resulting OECD (2007) report on Principles and Guidelines for Access to Research Data from Public Funding. Arzberger et al. (2004, 135) define data as "the raw material at the heart of the scientific process". The OECD (2007, 14) report characterizes data as:

factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

These definitions support a particular epistemic understanding, in which data are treated a priori as decontextualized bounded objects that carry inherent

meanings. In the words of Knorr-Cetina (2001, 191), data are understood as “closed boxes” that one views from the outside “as one would look at tools or goods that are ready to hand or to be traded further”.

From this perspective, data are perceived as separate, rather than inseparable, from the contexts that generate them. While data may be enhanced through the provision of contextual information and metadata, fundamentally, context is understood as conferring additive rather than constitutive meaning. That is, ontologically, the relationship between object and subject, or data and context, is understood in terms of separation rather than relationality. The ontological unit is taken to be “independent objects with independently determinate boundaries and properties” (Barad 2007, 33) rather than ontologically inseparable and entangled relations. This implies a particular model of knowledge-creation or epistemic practice. In Ingold’s (2010) terms, this epistemic practice is “hylomorphic”: form (morphe) is imposed upon passive matter (hyle). In a scientific context, this means that researchers, with particular goals in mind, act upon inert data or objects.

These epistemic assumptions, however, appear at odds with the epistemic practices of many natural and social scientists. Studies suggest that while understandings of what constitutes data vary (e.g. Bowker 2005), many researchers characterize data, and their meaningfulness and utility, in context-dependent ways (e.g. Wallis et al. 2007). Some recognize different forms of data that range across a continuum from “raw data” (decontextualized data) to “interpreted data” (contextualized data). Several suggest that “raw data” have limited scientific value when taken out of their contexts of production, signalling that meaning is not inherent to the data but to the data-in-context (e.g. Zimmerman 2008). Related to this, studies of scientific data sharing in the natural sciences suggest that researchers do not view data generation as a goal in itself; rather, they see it as tied to the specific scientific questions, contexts and projects they are working on (e.g. RIN 2009).

More generally, studies of knowledge-creation and epistemic practices challenge the assumed boundedness of data, as well as the separation between object and subject. Ingold (2010, 92) argues that all forms of “making”, including knowledge creation, are best understood in terms of “an ontology that assigns primacy to the processes of formation as against their final products, and to the flows and transformations of materials as against states of matter.” Knorr-Cetina (2001, 191) notes that epistemic objects (data) are defined by their capacity to unfold indefinitely, and their “lack of completeness of being that takes away their wholeness, solidity, and the thing-like character they have in our everyday conception”. She also suggests that epistemic practices are relational processes; that in carrying out their scientific work, researchers draw upon and deploy “relational resources” that are similar to the “being-in-relation” of everyday life. These resources include: taking the role or perspective of the other, where the other can be a protein, a bacteria, a tree or a person; making an emotional investment, or taking an interest in, the other; exhibiting moral solidarity and altruistic behaviour that serves the other; and imagining the other’s emotions, and engaging in a form of empathy. She

argues that it is through these relational practices that researchers reap insights, clarity and “epistemic dividends” (Knorr-Cetina 2001, 189).

The Moral Framework Underlying Open Access Data Sharing Policies

The moral case for open access data sharing draws on **utilitarian** (a particular form of consequentialism) and **deontological** theories. **Consequentialism** holds that **actions are morally right if they produce good outcomes**, and utilitarianism specifies that **these outcomes be useful** (Copp 2005). Deontological theories emphasize the importance of **independent moral rules and duties based on absolute principles about what constitute right and wrong courses of action** (Copp 2005). Both theories depersonalize moral actions by advocating universal principles that require moral agents **to make personal sacrifices for the general good** or in the fulfilment of their duty. Data sharing discourses draw on these moral theories in constructing data as a public good and data sharing as “good research practice” (Bishop 2009).⁵

Viewing data as “a public good” that belongs to everyone implies that they **therefore belong to no one in particular**. In practice, while researchers may agree that the public have a stake in publicly-funded data, they also understand data in personal and relational terms, as belonging morally to the researchers and respondents who generate them. Data constitute a form of personal intellectual property by virtue of their embodied production and the personal investments and relationships through which they are constituted. Investments of time, energy, effort, skill, emotion, intellect and physical labour mean that researchers become intimately entangled, inseparable from, and committed to the human and non-human “things” they study (Barad 2007; Knorr-Cetina 2001; Rees 2007; Zimmerman 2008). This explains why many social and natural scientists understand their data as part of their intellectual property; have strong emotional attachments to their data; feel they have a moral right to determine whether and how their data should be used, by whom, and for what purposes; and want recognition for the investments they have made in data collection. While the ethical requirement to obtain consent from research participants recognizes that human data belong (in part) to respondents, the notion that data may also belong to the researchers who generate them is rarely acknowledged, and is indeed contested (Baillie 2010).

Furthermore, while open access privileges scientists’ **moral obligations and responsibilities** towards the general public, scientists themselves tend to be more concerned about fulfilling their moral obligations and responsibilities towards specific respondents and co-researchers. Their practices tend to be guided, not so much by universalist moral principles, but by contextual, situational, relational, contingent and practice-based ethical approaches (see Edwards and Mauthner 2002) such as the “ethic of care” (Gilligan 1982). For example, researchers commonly uphold a **moral imperative to honour relationships of trust they have developed with respondents who have entrusted them with personal information and confidences** (whether DNA, medical histories, or personal narratives). They may use the data for the benefit of science and society, and to further their own careers,

but many will do so only in a context where they feel they can safeguard respondents' moral interests. This is why many researchers want to retain some personal control over their data, and over respondent protection. Open access policies, and their privileging of universal others, can be experienced as a violation of these personal and specific trust-based relationships and moral responsibilities. As Kaye et al. explain (2009, 333), "The obligation to share genomic data may be perceived as an imposition on the relationships that have been built between researchers and participants".

*The **Political** Framework Underlying Open Access Data Sharing Policies*

Unease with the assumptions underpinning the political agenda of open access may also fuel researchers' resistance to open access data sharing. The data sharing model which supports open access treats human and non-human data as interchangeable global commodities, which can be accessed internationally, selectively mined, retrieved and reassembled into new forms for new purposes, irrespective of their local, material, human and contextual specificities. Likewise, researchers are positioned as interchangeable data collectors who perform the technical task of amassing information, rather than individuals engaging in affective, material, and relational processes in the generation of data and knowledge. This overly cognitive model of data sharing obscures the politics of data production and the material power relations that exist between data producers and data users. Data sharing activities take place within a social and political context marked by power differentials between respondents, researchers and research agencies, and power relations amongst researchers. These hidden politics of open access may also account for obstacles to data sharing.

First, researchers may be concerned that open access policies exacerbate the power differentials and inequalities between researchers and respondents that have been problematized, particularly by feminist and post-colonial scholars (e.g. Ribbens and Edwards 1998; Wolf 1996). That is, there may be resistance to the normative ethical framework underlying open access whereby researchers are urged to seek respondents' blanket informed consent to universal and unconditional use of their data (whether anonymized or not) not only because it exposes respondents to risk and uncertainty, but also because it marginalizes respondents' moral and political rights to retain on-going involvement and decision-making powers in how their data will be used in the future.

Second, there may be resistance to the power relations between researchers and government-funded research agencies implicit within open access policies. Data sharing policies institutionalize a shift in control of research data. Funding bodies can require grant holders to deposit their data and institute punitive measures for non-compliance. In the case of the ESRC, for example, all award holders are required to offer their data for deposit and the ESRC will "withhold the final payment of an award if data have not been offered for archiving to the required standard within three months of the end of the award" (ESRC 2010, 4).



Furthermore, while individual researchers may in practice request a modification or waiver of the deposit requirements, and specify whether and how data should be archived, and with what levels of access (e.g. Bishop 2009), these requests are subject to approval from higher authorities. The ESRC, for example, reserves “the right to grant waivers only where sufficient evidence has been given demonstrating that data cannot be archived” (ESRC 2010, 4). By institutionalizing a shift in the definition and control of data and knowledge production away from researchers towards funding agencies, data depositories or governments, open access policies erode the authority and autonomy of scientists to decide whether and how to archive and/or share their data. This sits uncomfortably with researchers who want to retain control over data sharing and reuse; who want the discretion to “choose what to share, with whom, and when” (RIN 2009, 5; see also Perry 2008); and who favour consultative arrangements, direct contact and discussions with potential users, and autonomy to decide whether and when data should be released for specific reuses. A recent example of how the control of research data is changing is the case of Mike Baillie, an ecologist from Queen’s University Belfast. In April 2010, Baillie was forced to release tree-ring data under the Freedom of Information Act. The Information Commissioner’s Office ruled that Queen’s University Belfast must release the data to the public because Baillie did all the work while employed at a public university. Baillie, however, claimed that the tree-ring data he had collected over a 40-year period were his own personal intellectual property (Baillie 2010).

Third, open access data sharing models, policies and discourses also obscure the politics of knowledge production within research teams, and power relations between researchers. In practice, most data are produced by junior researchers, PhD students and/or technicians who, because of their structural positions and/or career stage, may lack the time and resources that senior researchers have to make full and timely use of the data they generate. While their data collection efforts are usually, though not necessarily, recognized and rewarded within the team through, for example, joint publications, recognition of their labour is less likely within the context of open access data sharing as currently constituted. Power and status differentials between team members risk leaving junior researchers open to exploitation as principal investigators have responsibility for making decisions about data sharing. Even where senior colleagues consult junior team members, by dint of their power and status they may privilege their own objectives. Within the context of international projects, there is a risk that data sharing becomes a form of scientific neo-colonialism. While open access potentially provides postcolonial contexts with easy and cheap access to data generated elsewhere, they may lack the necessary scientific, technical, digital or cultural resources to make full and speedy use of the data (see Luo and Olson 2008). In practice, it may be primarily well-resourced researchers and nations who stand to gain from open access, by reaping benefits and rewards from data generated by less well-resourced researchers and nations. From this perspective, the global data sharing project risks reproducing exploitative relations between nations, and between data users and data producers.

Rethinking Data Sharing Policies

The argument developed in this paper is that limited compliance amongst many academic communities with recently-introduced open access data sharing policies may stem, not so much from problems with digital data sharing infrastructure, but more from the failure of these policies to recognize data sharing as a relational practice. In the natural sciences, for example, the long established ethic, culture and tradition of data sharing is based on informal and personal contacts, relationships of trust and reciprocity, and data exchange on an individual basis through researcher networks (Wallis et al. 2007; Zimmerman 2008). Researchers favour personal, consultative and collaborative approaches to data sharing, direct contact and discussions with potential users, and data sharing practices in which primary and secondary researchers work together and can share tacit and contextual knowledge of the data. These informal and personal data sharing activities are preferred over the use of large-scale centralized repositories (RIN 2009).

In contrast, open access data sharing policies embody an instrumental view of data in which data are seen as free-floating public commodities, openly available to anyone wishing to tap into their inherent meanings. Moreover, the human, local, social, emotional, intellectual, corporeal, moral, material and political contexts of data generation and knowledge creation are marginalized. This vision is at odds with researchers' own understandings of the centrality of relational practices in securing the epistemic, moral and political integrity of knowledge and its production. The material and relational entanglements researchers develop with the people and things they study enable us to understand the reasons why, and ways in which, researchers care about how their data are used, appropriated and represented, epistemically, morally and politically. Without this context and care, data sharing risks producing knowledge that may be limited in its scientific use and meaningfulness; violating the moral and political rights and interests of those directly involved in data generation and knowledge production; and generating exploitative and coercive relations amongst researchers, between researchers and respondents, and between researchers and government and other agencies that fund research.

If researchers are objecting to the scientific, ethical and political norms institutionalized by open access policies, alternative policies may be preferred that institutionalize a more relational approach to data sharing. These policies might usefully place greater emphasis on encouraging and facilitating, rather than expecting and requiring, researchers to archive research data (e.g. ARC 2009, 36). They might give researchers greater discretion and control over whether to deposit data for open access, as well as how to archive and share the data. This could be an inclusive and collaborative process with researchers working with research agencies (funders, journals, data depositories) and respondents. This model would involve developing appropriate data care and management strategies tailored to particular studies, the nature of the data, and the views and preferences of those involved in data generation and other stakeholders. It would represent a contextual and case-by-case approach to data archiving and sharing. In practice, these issues are

recognized by open access advocates, expressed within some data sharing policies (e.g. BBSRC 2010; MRC 2010), and put into practice by researchers as well as the archivists who care for and manage data depositories (e.g. Caplan 2008; Henderson 2009; Mort et al. 2006). Nevertheless, many data sharing policies and guidelines remain highly prescriptive and fail to institutionalize, as normative, researcher and respondent autonomy and discretion, and multiplicity in data sharing models (e.g. CIHR 2007; ERC 2007; ESRC 2010).

A good example of an alternative model is the *Mass Observation Archive* and guidance notes, which implicitly recognizes data sharing as a relational practice. The guidelines convey a strong ethic of care in their management of archival documents, respect for those who have generated and deposited the papers in question, recognition of the emotional challenges of “letting go” of personal documents, and endorsement of a multiplicity of views and preferences with regard to document archiving and sharing. For example, the emotional attachments people form with their documents and data-sets are recognized and potential depositors are asked to “Make sure that you are really ready to part with your collection” because they “will probably have a sentimental value for you or for other members of your family” (Sheridan 2008, 1). In specifying a preference for “unedited material written at the time” (Sheridan 2008, 1), the archive implicitly privileges the epistemic integrity of the documents over their immediate availability. Moreover, while the archive indicates a preference for open access, depositors’ wishes are positioned as the prime concern: “My preference is for all the material we accept to be as open as possible, but of course, we would respect any special conditions which seem necessary” (Sheridan 2008, 2). Similarly, depositors’ preferences with regard to their involvement in third party use of their documents, whether they wish the material to be used in an anonymized way, and whether they wish to be named and credited are all acknowledged. Compared to many of the research funding agency data sharing policies and associated guidelines, the *Mass Observation* guidance notes are distinct in their tone, content and priorities.

Institutionalizing relational approaches to data sharing, and respondent and researcher consultation and involvement in data sharing, as normative principles may address many of researchers’ concerns by giving them greater autonomy and discretion over data archiving and sharing. This alternative approach would signal institutional recognition of the human subjectivities and relationships through which data and knowledge are produced. Notwithstanding its attractiveness, it is acknowledged that such a model is likely to entail time-consuming and resource-demanding practices that sit uneasily with the economic imperative underlying the open access movement. A significant driver for the current move towards data archiving is economic in that it allows the state to realize a return on its financial investments. As Arzberger et al. (2004, 135) make clear, the goal of open access is “to ensure that both researchers and the public receive optimum returns on the public investments in research”. That is, the economic case for data sharing being efficient and cost-effective rests in large part on a model of data sharing that has been stripped of the messy and complex human relationships through which data

and knowledge are produced. Yet, it seems likely that until data sharing policies recognize these relational complexities, institutionalize practices for addressing them, and factor in their resource implications, the global data sharing project will continue to meet resistance across many science communities.

Notes

- [1] The more general principle of free access to archives, however, dates back to the 1966 Archival Congress, held in Washington and convened by the International Council on Archives. For the first time, an international principle of free access to archives was declared and this led to the liberalization of access across democratic countries (Valge and Kibal 2007).
- [2] The term “open data” is also increasingly being used to refer to open access to research data in science (and government).
- [3] The concept of open access to research data dates back to the 1950s, when it was institutionally established through the formation of World Data Centers (particularly in the geo-physical sciences) designed to minimize the risk of data loss and maximize data access (Shapley and Hart 1982).
- [4] Cauldfield et al. (2008) question policies that advocate unrestricted data sharing of genome data because of privacy risks associated with public access to genomic information. They recommend releasing genome data into databases which restrict access to bona fide researchers and provide some phenotypic information linked to the genotypic data, increasing the scientific utility of the data.
- [5] It is worth pointing out that this represents a shift in institutional understandings of the moral status of research data in that, until recently, human data were understood a priori in personal rather than public terms, and data management policies included the destruction of human data within a specified time period (Willinson 1998).

References

- Arzberger, P., P. Schroeder, A. Beaulieu, G. Bowker, K. Casey, L. Laaksonen, D. Moorman, P. Uhler, and P. Wouters. 2004. Promoting access to public research data for scientific, economic, and social development. *Data Science Journal* 3: 135–52.
- Australian Research Council (ARC). 2009. Discovery Projects: Funding Rules for Funding Commencing in 2011 [cited 24 March 2010]. Available from http://www.arc.gov.au/pdf/DP11_fundingrules.pdf; INTERNET.
- Baillie, M. 2010. Tree-ring patterns are intellectual property, not climate data. *The Guardian*, Tuesday 11 May.
- Barad, K. 2007. *Meeting the universe halfway: Quantum physics and the entanglement of matter and meaning*. Durham: Duke University Press.
- BBSRC. 2010. Data sharing policy [cited 12 March 2010]. Available from http://www.bbsrc.ac.uk/web/FILES/Policies/data_sharing_policy.pdf; INTERNET.
- Benson, D. A., I. Karsch-Mizrachi, D. J. Lipman, J. Ostell, and D. L. Wheeler. 2008. GenBank. *Nucleic Acids Research*, 36. Database issue D25–30.
- Bishop, L. 2005. Protecting respondents and enabling data sharing: Reply to Parry & Mauthner. *Sociology* 39 (2): 333–6.
- . 2006. A proposal for archiving context for secondary analysis. *Methodological Innovations Online* 1 (2) [cited 11 July 2009]. Available from <http://sirius.soc.plymouth.ac.uk/~andyp/viewarticle.php?id=26>; INTERNET.
- . 2009. Ethical sharing and reuse of qualitative data. *Australian Journal of Social Issues* 44 (3): 255–72.

- Bowker, G. C. 2005. *Memory practices in the sciences*. Cambridge, MA: MIT Press.
- Broom, A., L. Cheshire, and M. Emmison. 2009. Qualitative researchers' understandings of their practice and the implications for data archiving and sharing. *Sociology* 43 (6): 1163–80.
- Bruna, E. M. 2010. Scientific journals can advance tropical biology and conservation by requiring data archiving. *Biotropica* 42 (4): 399–401.
- Budapest Open Access Initiative (BOAI). 2010. Available from <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm> (accessed 8 February 2010).
- Cambon-Thomsen, A., E. Rial-Sebbag, and B. M. Knoppers. 2007. Trends in ethical and legal frameworks for the use of human biobanks. *European Respiratory Journal* 30: 373–82.
- Canadian Association of University Libraries (CARL). 2009. Research data: Unseen opportunities [cited 25 February 2010]. Available from http://www.carl-abrc.ca/about/working_groups/pdf/data_mgt_toolkit.pdf; INTERNET.
- Caplan, P. 2008. Concepts of healthy eating: The history of two ESRC-funded research projects on food in the 1990s [cited 13 February 2010]. Available from <http://www.data-archive.ac.uk/doc/5801%5Cmrdoc%5Cpdf%5C5801uguide3.pdf>; INTERNET.
- Cauldfield, T., et al. 2008. Research ethics recommendations for whole-genome research: Consensus statement. *PLOS Biology* 6 (3): 430–5.
- Canadian Institutes of Health Research (CIHR). 2007. Policy on access to research outputs. Available from http://www.cihr-irsc.gc.ca/e/documents/cihr_policy_on_access_to_research_outputs_e.pdf (accessed 23 August 2010).
- Copp, D. 2005. *The Oxford handbook of ethical theory*. Oxford: Oxford University Press.
- Couzin, J. 2008. Genetic privacy: Whole-genome data not anonymous, challenging assumptions. *Science* 5 (321/5894): 1278.
- Dicks, B., B. Mason, M. Williams, and A. Coffey. 2006. Ethnography and data reuse: issues of context and hypertext. *Methodological Innovations Online* 1 (2). Available from http://erd.plymouth.ac.uk/mionline/public_html/viewarticle.php?id=28&layout=html (accessed 10 July 2011).
- Digital Curation Centre (DCC). 2010. Cross council policy overview. Available from <http://www.dcc.ac.uk> (accessed 19 July 2010).
- Edwards, Rosalind and Mauthner, Melanie. 2002. Ethics and feminist research: Theory and practice. In *Ethics in Qualitative Research*, edited by M. Mauthner, M. Birch, J. Jessop and T. Miller, pp. 14–31. London: Sage.
- European Research Council (ERC). 2007. ERC scientific council guidelines for open access [cited 25 March 2010]. Available from http://erc.europa.eu/pdf/ScC_Guidelines_Open_Access_revised_Dec07_FINAL.pdf; INTERNET.
- Economic and Social Research Council (ESRC). 2008. Memorandum from the Economic and Social Research Council in Response to the Ministry of Justice Consultation on the Use and Sharing of Personal Information in the Public and Private Sectors. Available from <http://www.justice.gov.uk/docs/Economic-and-Social-Research-Council-Data-Sharing-Response.pdf> (accessed 15 August 2009).
- . 2010. ESRC Research data policy [cited 30 November 2010]. Available from http://www.esrc.ac.uk/ESRCInfoCentre/Images/ESRC_Research_Data_Policy_2010_tcm6-37350.pdf; INTERNET.
- Gardner et al. 2003. Towards effective and rewarding data sharing. *Neuroinformatics* 3: 289–95.
- Gilligan, C. 1982. *In a different voice*. Cambridge, MA: Harvard University Press.
- Greely, H. T. 2007. The uneasy ethical and legal underpinnings of large-scale “Genomic Bio-Banks”. *The Annual Review of Genomics and Human Genetic* 8: 343–64.
- Griffin, C. 1997. Archiving qualitative datasets: Social research and the politics of control. *Radical Statistics* 66: 44–7.
- Hammersley, M. 1997. Qualitative data archiving: Some reflections on its prospects and problems. *Sociology* 31 (1): 131–42.

- Heaton, J. 2004. *Reworking qualitative data*. London: Sage.
- Henderson, S. 2009. Preparing qualitative longitudinal data for sharing and re-use: Making the long view lessons for timescapes 2009 [cited 13 February 2010]. Available from http://www.timescapes.leeds.ac.uk/assets/files/timescapes/MLV_Lessons.pdf; INTERNET.
- Hine, C. 2006. Databases as scientific instruments and their role in the ordering of scientific work. *Social Studies of Science* 36: 269–98.
- Ingold, T. 2010. The textility of making. *Cambridge Journal of Economics* 34 (1): 91–102.
- Kaye, J., C. Heeney, N. Hawkins, J. de Vries, and P. Boddington. 2009. Data sharing in genomics—re-shaping scientific practice. *Nature Reviews Genetics* 10: 331–5.
- Knorr-Cetina, Karin. 2001. Objectual practice. In *The practice turn in contemporary theory*, edited by T.R. Schatzki, K. Knorr-Cetina, and E. Savigny, pp. 184–97. Florence, KY: Routledge.
- Luo, Airong, and Judith S. Olson. 2008. How laboratories affect scientists from developing countries. In *Scientific collaboration on the Internet*, edited by G. Olson, A. Zimmerman, A. and N. Bos, pp. 365–76. Cambridge, MA: MIT Press.
- McLeod, J., and R. Thomson. 2009. *Researching social change: Qualitative approaches*. London: Sage.
- Medical Research Council (MRC). 2010. MRC Policy on data sharing and preservation [cited 9 March 2010]. Available from <http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Datasharinginitiative/Policy/index.htm>; INTERNET.
- Moore, N. 2007. (Re)using qualitative data. *Sociological Research Online* 12 (3) [cited 15 March 2009]. Available from <http://www.socresonline.org.uk/home.html>; INTERNET.
- Morrow, V. 2009. The ethics of social research with children and families in young lives: Practical experiences. Working Paper No. 53, Department of International Development, University of Oxford.
- Mort, M., I. Convery, C. Bailey, and J. Baxter. 2006. The health and social consequences of the 2001 foot & mouth disease epidemic in North Cumbria. Archiving the study data background guide for users [cited 13 February 2010]. Available from <http://www.data-archive.ac.uk/doc/5407/mrdoc/pdf/q5407userguide.pdf>; INTERNET.
- National Research Council (NRC), Committee on Issues in the Transborder Flow of Scientific Data. 1997. *Bits of power: Issues in global access to scientific data*. Washington, DC: The National Academies Press.
- Nelson, B. 2009. Data sharing: Empty archives. *Nature* 461: 160–3.
- O'Brien, S. J. 2009. Stewardship of human biospecimens, DNA, genotype, and clinical data in the GWAS era. *The Annual Review of Genomics and Human Genetics* 10: 193–209.
- OECD. 1994. *Global change of planet earth*. Paris: OECD.
- . 2007. *Principles and guidelines for access to research data from public funding*. Paris: OECD. [cited 16 March 2009]. Available from <http://www.oecd.org/dataoecd/9/61/38500813.pdf>; INTERNET.
- Office for Science and Innovation (OSI). 2007. Developing the UK's e-infrastructure for Science and Innovation. Office for Science and Innovation—OSI e-infrastructure Working Group [cited 16 March 2009]. Available from <http://www.nesc.ac.uk/documents/OSI/report.pdf>; INTERNET.
- Office of Science and Technology Policy (OSTP). 1991. *Policy statements on data management for global change research*. Washington, DC. DOE/EP-0001P Available from <http://www.gcric.org/USGCRP/DataPolicy.html> (accessed 28 January 2013).
- Open Access Directory (OAD). 2010 Timeline of the open access movement. Available from <http://oad.simmons.edu/oadwiki/Timeline> (accessed 10 February 2010).
- Parker, C. A. 2007. Institutional repositories and the principle of open access. *New Mexico Law Review* 37: 1–46.
- Perry, C. M. 2008. Archiving of publicly funded research data: A survey of Canadian researchers. *Government Information Quarterly* 25: 133–48.

- Rees, A. 2007. Reflections on the field: Primatology, popular science and the politics of personhood. *Social Studies of Science* 37: 881–907.
- Richardson, J. C., and B. S. Godfrey. 2003. Towards ethical practice in the use of archived transcribed interviews. *International Journal of Social Research Methodology* 6 (4): 347–55.
- Rin. 2009. *Patterns of information use and exchange: Case studies of researchers in the life sciences*. London: Research Information Network.
- . 2010. Stewardship of digital research data [cited 10 February 2010]. Available from <http://www.rin.ac.uk/our-work/data-management-and-curation/stewardship-digital-research-data-principles-and-guidelines>; INTERNET.
- Ribbens, J., and R. Edwards. 1998. *Feminist dilemmas in qualitative research: Private lives and public texts*. London: Sage.
- Ruusalepp, R. 2008. A comparative study of international approaches to enabling the sharing of research data. Available from: http://www.jisc.ac.uk/media/documents/programmes/preservation/data_sharing_report_main_findings_final.pdf (accessed 17 July 2010).
- Shapley, A. H., and P. J. Hart. 1982. World data centres. *Eos, Transactions American Geophysical Union* 63 (30): 585–5.
- Securing a Hybrid Environment for Research Preservation and Access (SHERPA). 2009. Research funders' open access policies [cited 13 March 2010]. Available from <http://www.sherpa.ac.uk/juliet/>; INTERNET.
- Sheridan, D. 2008. Guidance on donating personal papers [cited 12 June 2009]. Available from http://www.massobs.org.uk/downloads/guidance_2008.pdf; INTERNET.
- Social Sciences and Humanities Research Council (SSHRC). 2009. Research data archiving policy [cited 1 August 2009]. Available from http://www.sshrc-crsh.gc.ca/site/about-crsh/policy-politiques/statements-enonces/edata-donnees_electroniques-eng.aspx; INTERNET.
- UKRDS. 2009. The data imperative: Managing the UK's research data for future use [cited 25 March 2009]. <http://www.ukrds.ac.uk/resources/>; INTERNET.
- Valge, J., and B. Kibal. 2007. Restrictions on access to archives and records in Europe: A history and the current situation. *Journal of the Society of Archivists* 28 (2): 193–214.
- Van den Berg, H. 2005. Reanalyzing qualitative interviews from different angles: The risk of decontextualization and other problems of sharing qualitative data. *Forum: Qualitative Social Research* 6 (1): Art. 30.
- Wallace, H. M. 2005. The development of UK Biobank: Excluding scientific controversy from ethical debate. *Critical Public Health* 15 (4): 323–33.
- Waller, M., and R. Sharpe. 2006. *Mind the gap: Assessing digital preservation needs in the UK*. York: Digital Preservation Coalition. [cited 26 March 2010]. Available from <http://www.dpconline.org/advocacy/mind-the-gap.html>; INTERNET.
- Wallis, Jillian C., Christine L. Borgman, Matthew S. Mayernik, Alberto Pepe, Nithya Ramathan, and Mark Hansen. 2007. Know Thy sensor: Trust, data quality, and data integrity in scientific digital libraries. In *European Conference on Digital Libraries 2007, LNCS*, vol. 4675, edited by L. Kovács, N. Fuhr and C. Meghini, pp. 380–91. Berlin: Springer-Verlag.
- Whitlock, M. C., M. A. McPeck, M. D. Rausher, L. Rieseberg, and A. J. Moore. 2010. Data archiving. *The American Naturalist* 175 (2): 145–6.
- Willinsky, J. 2006. *The access principle: The case for open access to research and scholarship*. Cambridge, MA: Massachusetts Institute of Technology.
- Willinson, D. J. 1998. Health services research and personal health information: Privacy concerns, new legislation and beyond. *Canadian Medical Association Journal* 159: 1378–80.
- Wolf, Dianne, ed. 1996. *Feminist dilemmas in fieldwork*. Boulder, CO: Westview Press.
- Zimmerman, A. S. 2007. Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries* 7 (1–2): 5–16.
- . 2008. New knowledge from old data: The role of standards in the sharing and reuse of ecological data. *Science, Technology and Human Values* 33 (5): 631–52.