DIGITAL DATA SHARING: A GENEALOGICAL AND PERFORMATIVE PERSPECTIVE

The fate of sociological qualitative data has emerged over recent years as a ‘matter of concern’ (Latour 2004) within the UK and beyond (Thomson et al 2014), with something of a moral panic about whether, where and how historical and contemporary datasets are being preserved; and if they are, whether in a state that makes them fit for reuse by new generations of scholars. This has spawned new fields of interest and investigation, with researchers debating, and engaging in, the archiving and secondary reuse of qualitative data; and grappling with stubborn methodological, ethical, moral, political and legal challenges. One response to these problems has been to try to fix them, all the while taking the phenomenon of data sharing itself as an ontological given. My own interest is in opening up this ontology to investigation: to take data sharing practices as object of inquiry. Drawing on an emerging body of work on ‘the social life of methods’ (Savage 2010, 2013, Ruppert et al 2013), and on theoretical resources from science and technology studies, and feminist studies of science, I take a genealogical and performative approach and ask: How, and with what effects, has data sharing come into being? What concepts and norms are embedded and enacted in the multiple practices that are constituting data sharing as a phenomenon? And can data sharing be (re)made differently?

Key words: data sharing; data archives; qualitative research; sociological archives; genealogy; performativity.

Encountering qualitative sociological archives

In 1997 a fellow sociologist and I wrote a grant application to the Economic and Social Research Council – the ESRC – the largest funder of social science research in the UK, for an ethnographic study of work and family life in the oil and gas industry in north east Scotland. Two years prior to this, the ESRC had introduced its datasets policy. This required researchers to justify why they had to collect new data rather than use existing sources; and if they were generating new data, they had to offer these data for archiving (see ESRC Research Data Policy 2010 for a current version of the policy). This was my first encounter with the question of creating and using sociological qualitative archives, and it raised questions for me that intensified during the course of – and long after – undertaking our research project.

One of the issues that troubled me was what I saw as an implicit assumption within the ESRC policy that data were stand-alone objects that carried inherent meanings, and that their ontological state remained unchanged as they moved from the primary research context, to the archive, to secondary reuse contexts. In the literature, this has come to be identified as the epistemological problem of data and how to adequately...
contextualize them; and although I admit to being one of the people who has contributed to framing this issue within these **epistemological** terms (see Mauthner et al 1998), I now see it as a more profound point about ontology: about the ontological assumptions that underpin data sharing policies and practices, and specifically the assumption that data are ontologically separate and separable from the multiple practices through which they are constituted (Mauthner 2012a; Mauthner and Parry 2013).

Another reason why this new policy exercised me is because it marked a dramatic break from previous practice. When I started out my career as a social scientist just a few years earlier in the late 1980s, it was seen as good practice to destroy the research data that we generated – particularly personal data – once these had been analysed and written up. And this was reflected in data protection policies adopted by Universities and funding agencies. Now, the notion of ‘good research practice’ was being radically, but uncritically, redefined. We were being asked to take it as given that what was seen as bad practice yesterday was, overnight, turned into good practice.

A third issue that I struggled with concerned the relationships that we have with our research respondents. As a doctoral and postdoctoral researcher I was trained within a feminist research tradition, which sought to foster non-hierarchical relationships with research respondents based on trust and reciprocity (a position that has since been problematized; see Doucet and Mauthner 2008). The new ESRC policy, however, seemed to be reconfiguring relationships between researchers and researched along new axes of power. It was not simply that power was being consolidated on the side of researchers; rather, it seemed to be vested in altogether new stakeholders: research funding agencies; universities; and the general public – for all these now also had a claim on the data generated through these research relationships; and this in turn played a part in reshaping the nature of these relationships. In this new context, it was going to become harder to build the kinds of relationships based on trust, confidentiality, privacy and protection that researchers like myself had been used to doing. And this challenged the practice of qualitative, and feminist, research that I and many others had been engaged in (Mauthner 2012b).

A final issue that I found problematic was this sense that something was being imposed on us: that we were being told how we must investigate; that it was being decided for us what constituted good research and good research practice; that what was being proposed was not neutral but carried a set of contingent and historically-specific ontological, epistemological and moral assumptions that were left unexamined; that these new norms were being naturalized and that we were being asked to simply take them as given; and that the regulation of research, and research methods, was therefore seen as an acceptable, unproblematic and neutral practice that had no effect on the very practice of research (Mauthner 2014).
Questions and challenges around qualitative sociological archives

I begin my paper with these introductory remarks by way of locating myself, and how I first came to the question of qualitative archives in the late 1990s; that is, as a feminist qualitative social scientist grappling with some challenges that were being raised by our funding council requirements (see Mauthner et al 1998; Parry and Mauthner 2004, 2005; Mauthner and Parry 2009).

Since then, a whole literature has emerged in which sociologists have been debating issues and challenges around the archiving and reuse of qualitative data – or what I want to call digital data sharing practices (e.g. Cheshire et al 2009; Holland et al 2006; Corden and Millar 2007a, 2007b; Thomson 2009; Thomson et al. 2003, 2014). As I read this burgeoning literature, both within sociology and wider afield within other disciplines (see Mauthner and Parry 2013), I see it grappling with three central questions.

What are data?

I sit on the Research Data Management Committee in my University, where we have been tasked with devising a data management policy for the institution, and the single most challenging question we have faced is how to define ‘data’. Not only is it proving impossible to define data in such a way that it speaks to all disciplines, but there is also a great deal of resistance amongst many university researchers to the idea or necessity of doing this in the first place: to coming up with a general definition that, by necessity, must distinguish what is to be counted as data from what will be called metadata or contextual information. This in turn raises challenging questions about the relationship between data and context; about whether and how data can be interpreted out of their contexts of production; and about how much context is enough to make meaningful interpretations of data?

What are the risks of sharing data?

A second question raised by emerging data sharing practices are their potential risks, with four kinds of risks being the focus of debate. (1) There are perceived potential risks to participants such as: How secure are the data? How meaningful is it for participants to give their informed consent to future unknown uses of their data? Should we anonymise data? Does anonymisation necessarily guarantee confidentiality, privacy and protection of respondents, particularly in a digital context where data-linkages may make absolute anonymity impossible? Do respondents have the right not to be anonymised? And if they are named, do researchers have responsibility for their digital footprints and the endurance of these over time? Does sharing data risk breaching relationships of trust with respondents? Is there a risk that regulating data sharing may lead to coercive relationship between researchers and respondents?

(2) Potential risks to researchers are also being highlighted. Is there a risk that researchers’ ideas will be scooped – published by others – if they share their data? Does data sharing leave researchers open to criticism and reputational damage if poor research practices are exposed or if their findings are proven wrong? Does data
sharing fail to adequately recognize, value and reward the personal and professional investments researchers make in collecting data? Does data sharing risk disclosing personal details that researchers have revealed during the conduct of interviews or fieldwork?

(3) Data sharing – and failure to do so – is also seen to raise potential risks for science generally. Does failure to share data risk scientific progress? Is there a risk that data will be misinterpreted in their reuse, and lead to erroneous findings? Does anonymisation affect the integrity of data and lessen their reuse value?

(4) And finally, there are perceived risks to society. If scientific progress is impeded, does this risk the well-being of society? If data sharing breaches relationships of trust with respondents, is there a risk this may generate lack of trust by public in researchers and/or an unwillingness to take part in research?

Who owns data?

A third set of questions, around the issue of data ownership, are being raised by emerging data sharing practices. Who has rights over data? Who should have rights over data: researchers, Universities, funding organization, the general public? How should questions of data ownership be settled - morally and/or legally? And by recourse to what legislation and what rights: intellectual property rights, freedom of information rights, data protection rights?

So how to think about these challenges? For the most part, they have been seen as technical obstacles that can be fixed by technical solutions such as:

• developing more sophisticated ethical protocols for data reuse that might address ethical concerns by ensuring better compliance with ethical guidelines (regarding informed consent and data anonymisation) and legal requirements for data sharing.

• formulating data standards and requiring that contextual information be archived alongside the data which in turn might help render ‘raw’ data more meaningful by increasing their representational accuracy.

• creating better incentives and rewards – such as treating the creation of a dataset as an output in its own right, much like a research publication – which might encourage researchers to share their data.

Underlying these technical fixes is the assumption that data sharing practices are, inherently, a good thing because they lead to better science and professional practice. From this perspective, data sharing practices are seen as an inevitable and necessary development to which obstacles must be overcome for the sake of science, progress, and the greater good. This is what we might call an ‘internalist’ story (Savage 2010): a story in which ontological assumptions and commitments are taken for granted, naturalized and left unexamined. And finally, this is also a story that sees itself as the only possible story; the only game in town.

While there is nothing wrong with this way of thinking about data sharing practices, it is just one of many possible approaches; and one that relies on specific ontological presuppositions in which data sharing practices are treated as an ontological given, and bestowed with inherent capacities and potential to transform science and
professional practice. This ontological assumption in turn makes it possible to tell a teleological (Savage 2010: 7) or developmental narrative, in which data sharing practices are seen to embody scientific progress and professionalization. So current approaches rely on implicit ontological commitments – commitments that are taken for granted and naturalized, rather than opened up for discussion and examination. Furthermore, one of the implicit ontological assumptions being made is that ontology itself is given; and that it is singular: there is a single reality, or multiple perspectives on a singular reality, rather than multiple realities (Law 2004, Mol 2002). These assumptions of ontological given-ness and singularity make it difficult to entertain the possibility of alternative ontological commitments, and therefore alternative conceptualisations of data sharing practices and the challenges that they are throwing up.

So to think about data sharing practices and challenges in a different way we need to draw on theoretical resources that start from a different ontological starting point. And here I want to turn to a tradition of work called the ‘social life of methods’ that is taking research methods and practices as ontological objects of study rather than ontological givens.

The Social Life of Methods

Over recent years a body of work has been emerging that has come to be identified as ‘the social life of methods’ (Savage 2013, Ruppert et al 2013), and that is informed by several theoretical influences including science and technology studies, actor network theory, material semiotics, and feminist studies of science. A number of sociologists have been studying methods – such as censuses, surveys, interviews, focus groups, ethnography, and emerging digital methods – to understand how they came into being and gained prominence; and what kinds of assumptions about individuals and society these methods implicitly relied on and in turn materialised when they were put into practice (Law 2009a, Savage 2010). The point they make is that methods are not just neutral techniques for learning about a social world that is out there, and that we can describe, more or less, for practical purposes. To think of methods in this way and this way alone, they suggest, is to frame them in a particular and technical manner that misses out on the fact that methods also have a social life of their own. That is, methods do not “sit above the social world” but are themselves “embedded in contemporary life” (Savage 2010: 7). They are shaped by the social world in which they are located; and they in turn help to shape social realities.

To think of methods in this way is to think of them genealogically and performatively. First, we can look at the genealogy of methods and consider how they have come into being and taken on their present form, and the processes that have shaped these developments. Genealogical approaches refuse to see the rise of new methods in evolutionary or teleological terms. They denaturalize the present and imagine other possible futures.

Second, we can investigate the performativity of methods by examining how they shape and reconfigure the social world, and how they enact what Savage (2010) refers to as ‘unanticipated consequences’ or what Law (2009b) calls ‘collateral
realities’. From this perspective, methods are no longer understood as neutral or innocent techniques that put us in touch with a reality that is given, pre-existing, and already out there. Rather, as Law (2004) suggests, methods make realities. The processes of knowing the world – methods, their rules, and their practices – not only describe but also help to bring into being the realities they understand. That is, methods help to constitute realities that are generally taken as ontologically given and pre-existing. So within this approach, the ontology of the world is a question that is open to empirical investigation, a matter that has not already been decided.

**A genealogical and performative perspective on data sharing practices**

So what can we learn from bringing a genealogical and performative perspective to the question of digital data sharing practices and their challenges?

**Ontological becomings**

First, shifting away from the assumption that data sharing practices are an ontological given opens up questions about the genealogy or ontological becomings of these practices. For example we can ask: Why are data archiving, sharing reuse methods gaining increasing currency at this moment in time, within sociology and other disciplines? Through what processes and practices have these methods come to occupy a central place in what we might call the ‘social science apparatus”? How are these new practices shaping the development of sociology as a profession? What ethical norms are embedded and enacted in these practices? How have these ethical norms and frameworks taken shape? How did data sharing policies emerge, become normative, and displace data destruction policies? We can also ask genealogical questions about specific archives and their archival practices; about the qualitative research methods and practices through which data are constituted such as interview and transcription practices; and about data preparation and preservation practices, including data anonymisation and contextualization. We can also inquire into reuse practices including the technologies of data searching, and data analytic practices. And importantly, we can ask questions about the ontological assumptions embedded and enacted in these diverse practices; and how they have taken on their present ontological form. Thus, taking a genealogical approach opens up a whole new set of interesting questions about data sharing methods and practices.

**Ontological effects**

Second, seeing data sharing practices as more than just matters of technicality allows us to investigate what else these practices are doing *above and beyond* enabling us to archive and reuse data; that is, what their ontological effects are. We can examine how digital data sharing practices are reconstituting the very nature or ontology of research practices, relations and identities; and how they are helping to remake researchers, respondents, research relationships, ethical conduct, the public and the private, trust, the interview method, knowledge, social and political relations, and much more.
For example, qualitative researchers have become accustomed to the idea of disclosing details about themselves as part of building trust and rapport. It has also become standard practice with certain types of interviews to provide assurances about confidentiality and anonymity. For their part, respondents have become used to the idea that interviews are about disclosing personal and intimate life stories. Digital data sharing practices, however, may require establishing different kinds of relationships, that are less about trust between researcher and researched. And this in turn may be altering not only the stories that get told, but also what it is and means to be a researcher and to be a respondent. And it may also be that the nature of ethical conduct is in the process of being remade because while researchers have tended to see their ethical responsibilities as being primarily towards research respondents these responsibilities are being reconfigured to bring new stakeholders in, particularly society, the public, and the greater good.

Another example of the ontological effects of data sharing practices is the reconfiguration of ownership rights of research data: who has – or should have – rights; on what grounds rights can be claimed; and who gets to decide on these matters. It used to be taken as given that the labour of data collection constituted moral, if not legal, grounds for data ownership. We see this in that Universities have always had legal ownership of research data, but have, until recently, rarely exercised these rights with researchers thereby retaining their moral ownership rights. However, these moral rights are now being weakened by the regulation of data sharing as well as freedom of information legislation, which effectively confers data access and ownership rights onto the general public. So now, ownership rights can also be claimed by the general public, who fund research through their public taxes (see Mauthner 2014).

**Ontological politics**

Third, a focus on the ontological effects of data sharing practices brings into view what Mol (1999) calls ‘ontological politics’: the fact that these practices are not neutral or innocent but rather enact a politics, even though they tend not to recognise that they are doing so. We can investigate how data sharing practices are unsettling normative and implicit ontologies of research practices, relationships and identities and remaking them in specific ways.

One example of this is the way that data sharing practices require researchers to specify what are data and what is context. This, however, is not just a technical matter of deciding where to draw the line between data and context; or how much context is enough. It is also an ontological matter because preparing research data for archiving and reuse is an enactment of specific ontological assumptions. The normative practice of separating data from context – typically, separating interview transcripts which are seen as data from everything else which is seen as context – is just one way of archiving research materials. Another way might be to archive an entire project in all its messiness – fieldnotes, interviews, interview protocols, emails, recordings, minutes of project meetings, conference presentations, publications, grant applications, failed grant applications, associated projects, and so on – and call all of this ‘data’. Data preparation and archiving practices – as they have come to be
ontologically constituted – are neither neutral nor given. They are a specific way of enacting reality; just one of many possible ways of doing this.

Researchers are working not only with the ontological specificity of particular archives, and their ways of organizing materials, but also within institutional contexts that also specify certain definitions of data and context. So for example in the UK funding organisations are informed by the Organisation for Economic Co-operation and Development’s (2007) *Principles and Guidelines for Access to Research Data from Public Funding* and its vision for data sharing as well as its definition of data, which is:

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.

This definition of data reflects a very specific set of ontological assumptions in which ‘data’ are taken as the basic unit of analysis and key building block of the scientific enterprise. But it is possible to imagine alternative configurations in which, for example, it is knowledge practices – that is what researchers actually do vs what they produce – that are taken as the primary ontological building blocks.

So approaching the issue of archiving, sharing and reusing research data from a social life of methods perspective brings into the focus the politics of data sharing practices by highlighting the power of the mundane instruments – such as archives and methods – which are actually deployed in research activity, and the power relations which are enacted through the institutional embedding of these practices in data sharing policies.

**Imagining other possible futures**

As I see it, the problem is not with data sharing methods per se but rather with the normativities and moralisms that are attached to them (Law 2004): the presumption that we should be engaging in these practices and that these practices inherently constitute good research and good research practice. The problem is that this position relies on a very specific set of ontological assumptions, including the assumption that ontology itself is a given and that it is singular. These combined assumptions make it very difficult to debate ontological issues, or entertain the possibility of there being other ontologies. And I would argue that it is these more deep rooted ontological issues that underlie what Borgman (2012) refers to as the conundrum or ‘intricate and difficult problem’ of sharing research data – the reality that, in her words, “very little data is being shared, despite the best efforts of funding agencies and journals”; as well as the empty archives problem that Nelson (2009) discusses whereby despite enormous investments in the creation of digital repositories, and the fact that most researchers agree that open access to data is the scientific ideal, in practice, many researchers choose not to share.

My own interest is in exploring how taking up a different ontological starting point, and seeing data sharing practices and their challenges through a genealogical
and performative perspective, might suggest other possible practices and ways of doing things. If we see current practices as neither necessary nor inevitable, then it becomes possible to think about other ways of building archives; preserving, curating and revisiting research; engaging in research relationships and ethical conduct; being a responsible researcher and respondent; and much more. And in considering whether things can be made or done otherwise, we can look at what is at stake in reconfiguring the ontology of these practices, relations and entities; at what constituencies stand to gain or lose from these shifts. We can begin to have a conversation about the ontological politics of data sharing practices, and to imagine other possible futures.

References


Mauthner, Natasha S., Parry, Odette. and Backett-Milburn, Kathryn. 1998. The data are out there, or are they? Implications for archiving and revisiting qualitative data. *Sociology,* 32 (4), 733–745.


Użyczanie danych cyfrowych: perspektywa genealogiczna i performatywna

Streszczenie

Dalsze losy socjologicznych danych jakościowych stał się w ostatnich latach w Wielkiej Brytanii i poza nią „niepokojącą kwestią”, przejawiającą się w formie, która umożliwiłaby ich ponowne wykorzystanie przez nowe pokolenia naukowców. Przyczyniło się to do powstania nowych obszarów zainteresowań i dociekań badaczy dyskutujących i angażujących się w archiwizowanie oraz ponowne wykorzystywanie danych jakościowych i zmagających się z uporządkowanym, nieodzownym kwestiami metodologicznymi, etycznymi, moralnymi, politycznymi i prawnymi. Dotychczasowe próby rozwiązania tych problemów były na zakończeniu, że zjawisko dzielenia się danymi jest ontologicznym pewnikiem. Ja natomiast chciałabym zakwestionować tę perspektywę, potraktować praktyki dzielenia się danymi jako coś nieoczywistego i wartego dociekań. Odsuwając się od wyrastającego korpusu prac na temat „społecznych zjawisk” (Savage 2010, 2013, Ruppert i in 2013) oraz do teoretycznych inspiracji ze studiów nad nauką i technologią a także feministycznych studiów nad nauką przyjmuję podejście genealogiczne i performatywne i pytam: W jaki sposób doszło do głosu idei „dzielenia się” danymi i jakie są jej konsekwencje? Jakie pojęcia i normy wpisane są i aktywowane w rozmaitych praktykach archiwizowania i udostępniania danych? I czy „dzielenie się” danymi da się pomyśleć i praktykować inaczej niż dotąd?

Główne pojęcia: użycie danych/dzielenie się danymi; archiwa danych; badania jakościowe; archiwa społeczne; genealogia; performatywność.