

Archival practices and the making of memories

Natasha Mauthner
University of Aberdeen
Business School
Dunbar Street, Old Aberdeen
n.mauthner@abdn.ac.uk

Judit Gárdos
Hungarian Academy of Sciences
Centre for Social Sciences
Országház utca 30, Budapest
gardos.judit@tk.mta.hu

ABSTRACT

Discussions about social science digital archives have tended to address methodological, technological, ethical and legal issues such as what digital tools to use; what metadata to collect; how to curate numerical or interview data; and how to make archived materials available in ethical and legal ways. These discussions take place on the assumption that these archival practices are ontologically independent from archival materials. Following Derrida (1995), we want to explore the relationship between archival practices and archival documents on the assumption that ‘archivization produces as much as it records the event’ (Derrida 1995:17). On this approach, archival practices are understood as non-innocent, culturally and historically-specific practices that, in the act of ‘preservation’, help *make* specific ‘memories’ at the expense of others (Barad 2007, Derrida 1995, Foucault 1972). In this paper we take up this issue in relation to the curation of social science quantitative research data. We conceptualise data curation practices in broad terms as including specific practices—e.g. cleaning up of datasets; data anonymisation; streamlining interviews; data storing, categorising and visualisation; data search tools; etc.—as well as a wider range of knowledge, ethical, legal, political and economic practices these practices are entangled with—e.g. field- and discipline-specific knowledge-making practices; national/international data management and curation policies, practices and guidelines; ethical guidelines provided by professional bodies; country-specific data protection, copyright, and information sharing legislation; etc. For the purposes of this paper we focus on three specific data curation practices—data cleaning, data anonymisation and metadata preparation—and investigate the ontological processes through which these practices help constitute the survey data they ostensibly archive.

Keywords

archival theory, archival practices, social scientific knowledge production, data curation

1. INTRODUCTION

In both archival and social sciences data curation practices have tended to be understood in technical and administrative terms: as neutral means of safeguarding documents and data, where objectivity is secured by following agreed professional standards and protocols. Postmodern thinking coupled with technological change have challenged this normative understanding of the archive. The work of Derrida (1995) and Foucault (1972) has been particularly influential and has caused archivists and scholars to reconfigure archival science and practice to explore: the nature of archives as socially constructed institutions, the role of archives in the production of knowledge about the past, and the power of archives and

records to shape our notions of history, identity and memory (Schwartz and Cook 2002, Cook and Schwartz 2002).

In this paper we take up these ideas specifically in relation to the curation of social science quantitative research data. Our argument is that data curation practices do not innocently preserve data but help constitute their ontological identity. Data curation practices are ‘performative’ in that they help bring into being the data they ostensibly preserve (Butler 1990, Barad 2007, Law 2004). They do so, we suggest, not simply through the subjectivity of data curators or the technologies of the data repositories but rather through the specific conceptual assumptions that are embedded and enacted in data curation practices. Data curation and archival practices, we suggest, can be understood as historically and culturally-specific and contingent ‘metaphysical practices that necessarily enact specific metaphysical commitments to the exclusion of others’ (Mauthner 2015:331). On this approach, data curation practices (along with wider sets of knowledge, legal, technological, political, economic, ethical, and moral practices) are an ineliminable and constitutive part of the data they help bring into.

Our paper is organized in five parts. We begin with a discussion of Derrida’s *Archive Fever*, and go on to consider how his ideas, and postmodern scholarship more generally, have contributed to discussions about the philosophy of the archive in archival science. We then turn to consider the nature and extent of philosophical debate in relation to social science data archives. In the next section we outline some theoretical ideas around the performativity of knowledge-making practices that inform our own attempts to conceptualise data curation practices in performative/non-representational terms. In the final part of the paper we provide some examples of how data curation practices help *make* the nature of data by arguing that these practices embed and enact (reproduce) specific historically- and culturally-contingent concepts and categories.

2. JACQUES DERRIDA’S ARCHIVE FEVER

In *Archive Fever*, Jacques Derrida (1995) takes the archive as his object of study. He challenges normative understandings of the archive as a neutral and innocent keeper of historical records and artifacts that tell of a past that really happened, and that can be returned to for the purposes of constructing originary tales. Derrida troubles this commonsensical view by reminding readers of the etymology of the word ‘archive’. The term, he suggests, refers to the *arkhē* – the source, beginning, origin - in two senses: ontological and nomological. *Arkhe* denotes at once the commencement and the commandment. The meaning of ‘archive’, Derrida goes on, comes from the Latin

archivum or *archium*, itself derived from the Greek *arkheion*. The latter was the residence of the *archons*: the superior magistrates who commanded and ruled in ancient Greece. On account of their publicly recognized authority, their home became the repository of official documents. Not only did the archons become the guardians of these documents, but they were also accorded hermeneutic rights: They had 'the power to interpret the archives' (Derrida 1995:3).

Derrida points further to the politics of the very constitution – and not only interpretation - of archives and archival content. Archives come into being through political processes and practices. Their materialization is inseparably ontological and political. Archives, Derrida argues, come into existence in highly specific ontological and political configurations – what he calls 'a privileged *topology*' (Derrida 1995:3). It is in this sense that archives both *embody* and *enact* a specific politics: they are both expressions and instruments of power (Harris 2002). In Derrida's terms, archival 'documents in effect, speak the law: they recall the law and call on or impose the law' (Derrida 1995:3). The processes and practices through which archives are constituted both 'lay down the law and *give the order*' (Derrida 1995:7).

This order - in the form of the classificatory concepts and systems – is mutually constitutive of archives and archival records. As Derrida puts it: 'the technical structure of the *archiving* archive also determines the structure of the *archivable* content even in its very coming into existence and in its relationship to the future. The archivization produces as much as it records the event' (Derrida 1995:17). The systems through which archival content is organized/ordered – for example, whether a record is classified as 'theory' or 'private correspondence', 'biography' or 'autobiography' – contribute towards the *making* of archival content. Derrida calls this coming into existence in a specific place and according to a specific law (classificatory system) the 'archontic principle'. The archontic principle of the archive, he suggests, 'is also a principle of consignment, that is, of gathering together' (Derrida 1995:3) where '*Consignment* aims to coordinate a single corpus, in a system or a synchrony in which all the elements articulate a unity of an ideal configuration' (Derrida 1995:3).

Derrida's point is not that classificatory concepts and systems can or should be avoided as without them 'no archive would ever come into play or exist as such' (Derrida 1995:3). In order to exist, every archive must necessarily come into being in a specific form. The issue is rather that the politics and 'violence' (Derrida 1995:7) through which the archive must be, and is inevitably, constituted is naturalized and taken as given. Derrida wants to make apparent that the archive is a place 'where law and singularity intersect in *privilege*' (Derrida 1995:3). He argues that:

'A science of the archive must include the theory of this institutionalization, that is to say, the theory both of the law which begins by inscribing itself there and of the right which authorizes it. This right imposes or supposes a bundle of limits which have a history, a deconstructable history'.

Derrida points to the ways in which archives embody and enact power in two ways: first, through privileged topologies (concepts and categories) that help make the law (specific ontological and political realities); and second, by naturalizing,

forgetting, and erasing the fact that these *are* privileged topologies – that archives are a materialization of specific ontological and political commitments to the exclusion of all others. It is in this sense, Derrida argues, that the concept of the archive carries but also forgets the name *arkhē*.

3. THE PHILOSOPHY OF THE ARCHIVE

Derrida's writings on the archive, along with those of Foucault (1972) and other postmodern thinkers, have fuelled debate about 'the philosophy of the archive' (Whatley and Brown 2009). For over two decades, archivists and scholars have been grappling with 'the new realities being fashioned by technological revolution and postmodernist epistemologies' (Harris 2002:83), and their implications for the conceptualization and practice of archival work. Both postmodern thinking and the shift to electronic records have challenged normative and naturalized assumptions about the neutrality, objectivity and given-ness of archives, archival records, and archival practices. The shift to electronic record-making and keeping has helped make visible the hitherto invisible: the role of the archivist and archival practices in the very act of record-creation and preservation, thereby dispelling the notion of the archivist as innocent guardian of the archival record (Hedstrom 2002, Trace 2002). Similarly, postmodern scholarship has troubled normative technical understandings of archival science, rooted in 19th century positivist and administrative-judicial frameworks. As Cook (2013) notes, early literature on archiving (Jenkinson 1947) stresses that the goal was to "preserve the original records as evidence of the functional-structural context and actions that caused their creation". Cook (2013: 106) calls this approach pre-modern, the archivist being a passive keeper of an entire documentary (Cook 2013:102). Duff and Harris (2002) see the roots of this positivistic archival practice in the traditions of Enlightenment. According to them, the understanding behind these practices is this:

"The boundary between text and context is hard and stable. A record's context is bounded and readily knowable. The archivist's role in relation to records is to reveal their meaning and significance - not to participate in the construction of meanings - through the exercise of intellectual control." (Duff and Harris 2002:264)

Under the influence of the postmodern 'archival paradigm shift' (Cook 2001), archivists and scholars have challenged 'unquestioned assumptions underlying the nature and meaning of archives in society' (Schwartz and Cook 2002:5). They have destabilised representational approaches in which the archival record is seen to reflect or represent reality and 'stand for the facts it is about' (Duranti, 2001; see also MacNeil 2004). There has been a shift away from treating the archive-as-source to taking archives, archival records and archivists as objects of study in their own right (Stoler 2002).

The almost exclusive focus of archival science on the technology and mechanics of archival processes has given way to consideration of 'what archives, records, and archivists do on a philosophical or theoretical level, the power they wield, the impact they have' (Schwartz and Cook 2002:18-19; Cook and Schwartz 2002). A major focus of investigation has been on power (Cook 2001). Archivists and scholars have investigated how power is exercised both through the specificity of the cultural frameworks that shape archival records, archives as institutions and archival practices, and through the

naturalization of these implicit underpinning cultural frameworks. Archival records and archives are no longer viewed as naturally (given) constructs but as socially constructed. The research focus has shifted to the social, historical, cultural, and political power constellations, as well as to the standards, values and ideologies that shape the creation and maintenance of archives, archival records, and the archival profession (Trace 2002).

This postmodern archival practice, argues Cook, takes the form of social activism for “memory-meaning, adopting a flexible, fluid, and pluralistic mentalité mirroring the values of postmodern society and the possibilities of digital technology” (Cook 2013:111). In parallel to this, however, and bound to a modernist tradition of archiving, archivists are also constantly developing more sophisticated means by which archives are managed and evidence is protected – with a focus on creating and implementing standards, record-keeping requirements, process templates, and system architectures. Here, rigid consistency of professional practice is sought. Such work reveals, Cook (2013) suggests, the continuing concern for evidence among the memory-dominated and identity-formation paradigm. “As a result, between the poles of evidence and memory, there was sometimes considerable tension in professional discourses, between ever more sophisticated and complex modernist techniques for evidence protection reflecting a culminating expertise in that regard and ever more contextualized and contingent postmodern ideals in turn reflecting contemporary societal values” (Cook 2013:111). He sees the many contemporary archivists endeavours’ around electronic data preservation in the same old modernist line of tradition (Cook 2013:100-101) to preserve the data and how they came into being. As we do on to discuss, in the context of social scientific data archiving, such modernist practices can involve state-of-the-art metadata schemes, archival software products, etc.

4. THE PHILOSOPHY OF SOCIAL SCIENCE DATA ARCHIVES

As in the field of archival science, philosophical issues in relation to social science data archives have been discussed since the late 1990s but theoretical and methodological engagement and uptake of these ideas has been limited and little attention has been given to the philosophy and politics of social science data archives (Mauthner and Parry 2009, 2013, Mauthner 2015, Mottier 2005). Few systematic critical and/or ethnographic investigations have been undertaken exploring how the processes and practices through which social science projects and data are archived shape the resulting archived research collection. In contrast, there are many texts outlining normative data archiving and curating principles and practices and their underpinning ethical, legal, governance, scientific and technical frameworks and infrastructures. Poole (2015), for example, discusses the infrastructure of science data curation, including the roles of cyberinfrastructure, research communities, collaboration, planning, policy, and standards and best practices. He also addresses the role of institutions—archives, research libraries, institutional repositories, and centers—in curating science data and the role of archival principles, such as provenance, selection and appraisal, authenticity, metadata, risk management, and trust, in digital curation.

Overall, discussions about the archiving of social scientific data are rooted within ‘traditional archival thinking’ and a ‘solely technical framework’ (Trace 2002:150) that overlooks the processes of record creation and preservation. As Mauthner and

Parry (2013:56) suggest, ‘The focus has ... 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 (...). 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 (...)’.

The assumption within these approaches - that data can be treated independently of their ontological contexts of production - has been seen as particularly problematic for qualitative researchers influenced by the postmodern turn and working within interpretive and social constructivist traditions for whom data are understood as reflexively constituted through historically- and culturally-specific practices (Hammersley, 1997, 2010, Mauthner et al., 1998, Parry & Mauthner 2004, Savage, 2005). These philosophical concerns have been tackled through practices seeking to ensure that contextual information is archived alongside the data. This has been understood as rendering data (and resultant knowledge) more meaningful by enabling researchers to better understand the conditions through which data are generated (Corti 2011). The effect of these practices is to make possible the creation of a new object of study: data and the contexts that constitute them, rather than data alone (e.g. Savage 2005, 2010 Moore 2006, Thomson 2014). However, as Mauthner (2015) argues, the shift to taking context into account still conceptualises data archiving in the same technical terms in which data curating is understood as a neutral means of preserving data *and* context. These practices therefore continue to enact an ‘ontology of given realities’ (Mauthner 2015) – such as data and context - by overlooking the *constitutive* nature of data archiving and curation practices. Data and context are still taken to be ontologically separate and separable, rather than mutually constitutive.

5. A PERFORMATIVE CONCEPTUALIZATION OF KNOWLEDGE-MAKING PRACTICES

This enactment of data archiving and curation practices relies on largely hidden and commonsense *representational* assumptions about the nature of reality in which the world is understood as given, and the practices of knowing (e.g. preserving, archiving, curating) are bracketed out and treated as technique (Law 2004). Barad (2007:53) explains that representationalism ‘marks a failure to take account of the practices through which representations are produced’. A representational conceptualization of knowledge-making practices (e.g. data curation practices), she suggests, ‘takes the notion of separation as foundational. It separates the world into the ontologically disjunct domains of words and things, leaving itself with the dilemma of their linkage such that knowledge is possible’ (Barad, 2007:137). Drawing on Rouse (1996), Barad further argues that representationalism underpins both empirical realism and postmodern philosophical approaches that turn to language and discourse, as both share the representational belief that knowledge mediates access to the material world (reality). Where they differ is on what they take to be their referent: whereas realist claims are understood to represent things in the world as they really are (i.e. nature), social constructivist ones are seen to represent objects that are

the product of social activities (i.e. culture). Moreover, attempts to acknowledge the knower through reflexive approaches are also founded on representationalism in that they take for granted the notion that representations reflect (social or natural) reality. Reflexivity, Barad suggests, still holds the world at a distance: it 'is based on the belief that practices of representing have no effect on the objects of investigation and that we have a kind of access to representations that we don't have to the objects themselves' (Barad 2007:87).

On our account, current configurations of data curation practices, and the wider assemblage of practices they are part of, are being enacted on implicit representational terms because their constitutive role in the formation of data is overlooked. Our approach in this paper is to conceptualize data curation practices along nonrepresentational lines, through a philosophical framework that is able to materialize the constitutive effects of data curation (and wider) practices on their objects of study and knowledges produced. For this, we turn to Karen Barad's posthumanist performative metaphysics, a framework that embodies and enacts a non-classical ontology in which entities are not taken as given but as constituted through material-discursive practices. Barad's work challenges classical—Newtonian and Cartesian—metaphysics and provides a new ontology, epistemology, and ethics that she terms 'agential realism'. On her account, knowledge-making practices are an ineliminable and constitutive part of the realities they help bring into being. On our reading, her scholarship provides a distinctive metaphysical framework that can materialize, and help reconfigure, the representational ontological assumptions that are embedded and enacted in data curation practices and the wider regime of practices of which they are part.

Barad's framework is being taken up in the social sciences, with scholars developing different theoretical and methodological applications of agential realism including ways of approaching data archiving and revisiting (e.g. Tamboukou 2014). In this article we expand on Mauthner's approach to Barad's framework, and her broader programme of work that seeks to develop a nonrepresentational—posthumanist performative—conceptualization and enactment of research methods and knowledge-making practices in the social sciences. In particular, Mauthner argues that knowledge-making practices (such as data curation) make realities through the specific metaphysical and conceptual assumptions these practices embed and enact (see Mauthner 2015, forthcoming b, c).

6. THE PERFORMATIVITY OF SOCIAL SCIENCE DATA CURATION PRACTICES: SOME ILLUSTRATIVE EXAMPLES

In this section we explore data curation practices and the conceptual assumptions underpinning these practices. We draw on Judit Gardos' work at the Research Documentation Centre (RDC) of the Hungarian Academy of Sciences' Centre for Social Sciences where she works as part of a team of social scientists curating data from social science research projects. Their curation practices are informed by the Data Documentation Initiative-Lifecycle (DDI-L, <http://www.ddialliance.org/Specification/DDI-Lifecycle/>).¹ For

¹ The DDI-L is similar to the Digital Curation Centre's Curation Lifecycle Model (see Higgins 2008).

the purposes of this paper we focus our attention on the curation of social science survey data. We limit ourselves to three specific data curation practices: data cleaning, data anonymisation and metadata preparation. Following on from our discussion above of performative approaches to knowledge-making practices, our interest here is in exploring the conceptual assumptions underpinning these data curation practices.² Here, we build on insights from the field of archival science and the work of archivists and scholars, such as Schwartz and Cook (2002), who discuss the alleged value-free tools of archival practice - like standards and templates - and how these impose their own rational, systematic way of seeing on a world of records, record keeping and records creators (see also Trace, 2002, Hedstrom, 2002).

6.1 Process of curation

In the case of the Hungarian Research Documentation Centre archivists tend to consult - face-to-face or via email - the researchers who undertook the survey in an attempt to gather and (re)construct the meaning of specific variables and values. In this way, the nature and meaning of the data are constructed by researchers and archivists working together in an interactive practice of remembering, argumentation and guessing - and implicitly operating within a wider, historically and culturally specific, field of social science. This is akin to what Cook (2013:113-116) calls the community paradigm of contemporary archiving: the notion that archiving is transformed from an elitist practice - performed by the archivist - into a participatory practice embedded in the social scientific community and society more generally.

6.2 Data cleaning

Data cleaning involves tidying up survey values that appear not to make sense, are mislabelled, or are not provided. These practices are designed to address the problem of 'missing data', a term derived from statistical research which refers to instances where no data value is stored for the variable in an observation either because no data value was recorded or because the reader is unable to make sense of the data value provided. Some data values - such as 1964 or 1981 for 'year of birth' are self-explanatory. In other cases, however, the data values require explanation such as '1' and '2' in a variable labelled 'sex'. In the case of gender where the value '1' is labelled male and '2' is female we may find values of 3, 4, 9 or 99. In such cases, we usually convert these into missing values. The original values remain visible but will be excluded from statistical procedures. Following the DDI-L standard, and its aim to document every step of the data curation process, we add in the curation documentation how we proceeded with such values.

While such data cleaning practices are normatively understood as neutral and technical, on a performative approach they can be understood as helping to create data through, for example, the (re)assignment of meaning for specific values. The purpose of data cleaning practices is to make every value of every variable understandable for researchers - to create a dataset that is meaningful and usable by imagined and anticipated social scientists and social science research. Yet while there are technical guidelines for dealing with missing data - the UKDA for example informs data curators to look for "unlikely or impossible values for interval variables" - there is no

² Data are made by wide-ranging practices, including data collection and data analysis practices, and their underpinning conceptual assumptions. The focus of our paper, however, is specifically on data curation.

overarching paradigm in relation to what might be regarded as ‘unlikely’ or ‘impossible’ (Sana-Weinreb 2008:515). In practice, as Sana-Weinreb (2008) explains, notions of ‘likely’ or ‘possible’ are enacted in normative – culturally and historically specific – terms: for example, in most settings, a teacher is likely to have more than a primary school education; or ‘someone who reports no sexual contact in the last year is not likely to be young, currently married, and co-resident with their spouse’ (Sana-Weinreb 2008:517). It is therefore easy to imagine cases where data are not faulty but just reporting situations that are normatively understood as ‘unlikely’ or ‘impossible’. It is in this sense that, as we suggest, data cleaning practices embed and enact normative conceptual assumptions – in this case, about teacher training and sexual conduct.

Labelling practices are also informed by social science research practices more generally. For example, quantitative surveys and analysis of datasets typically classify income into 4 to 10 groups, which provides both anonymisation of the exact income level and enables standard ways to analyse data. Data curators normally have knowledge of how these income groups are constructed by the researchers. In cases where labelling is absent in a dataset, curators label the variables according to this prior knowledge. Their labelling practices therefore instantiate conceptual categories that are normative and accepted within quantitative social scientific work.

6.3 Data anonymisation

Anonymisation of research collections that contain personal data (i.e. data that can be connected with a single identifiable person) is a very common archival practice. There are various methods for anonymising data. Following the UKDA guidelines, the process of anonymisation depends on whether the data are quantitative or qualitative. According to the UKDA, the purpose of anonymising quantitative data is to protect the identity of respondents while retaining ‘as much meaningful information as possible’. Their guidelines suggest:

‘Aggregate or reduce the precision of a variable such as the respondent’s age and place of residence. As a general rule, report the lowest level of geo-referencing that will not potentially breach respondent confidentiality. The exact scale depends on the type of data collected, but very detailed geo-references like full postcodes, wards or names of small towns or villages are likely to be problematic.’³

As in the case of data cleaning practices, the question that arises is how ‘meaningful’ is being defined. In designating particular pieces of information for deletion or retention, data anonymisation practices therefore enact specific but implicit understandings of what is regarded as ‘meaningful’. These understandings are shaped by numerous influences: the data curators’ understanding of the research project in question; data curation practices more generally; the field of social sciences; the local, national, and international context, and much more. For example, in cases of datasets where there is a focus on higher education, information on whether a given city has a university or not may be deemed important and therefore retained; and this knowledge of whether the city has a university or not is only accessible through other sources such as prior knowledge of the university structure of a specific country. Another example, is where a research collection includes a focus on the Roma minority. In this case, it may be

deemed important and relevant to retain information on where the respondent lives and the extent of its Roma population.

These examples illustrate the ways in which data anonymisation is not a neutral practice deleting or retaining information as if the meaning of this information is given and self-evident. Rather, anonymisation practices enact specific normative historically and culturally-contingent definitions of what is ‘meaningful’, which come to be built into the very nature and fabric of the data.

6.4 Metadata preparation

Metadata are defined and understood as data that describe and give information about other data. They are seen to provide contextual information to facilitate ‘meaningful’ interpretation about the data and are viewed as being as important as the data themselves. To illustrate metadata we have randomly chosen a dataset on ‘Election Funding of Finnish MPs 2007’ accessed from the Consortium of European Social Science Data Archives’s (CESSDA) data catalogue [See Appendix 1]. On the left of the image a list of ‘variable descriptions’ is provided under ‘Metadata’. One of these, highlighted on the right of the image, is ‘The amount of contributions received from party organisations (in euros)’. Some basic statistics are also provided such as the number of ‘valid cases’. We make two observations about metadata preparation practices, their underlying assumptions, and their performative effects.

First, the categories and concepts of metadata – the variables listed such as candidate’s name, gender, year of birth, political party, etc. – are not neutral descriptors but rather help constitute data in specific ways according to classificatory systems that are currently normative, accepted and regarded as ‘meaningful’. Indeed, metadata schemes have been standardized (e.g. according to DDI-L scheme) because lack of standard formats is seen to hamper the ability to search across studies for similar items or similar studies (Rasmussen and Blank 2007:62). Through their classificatory systems, metadata practices act to institutionalise and further solidify these systems – systems which are specific ways not only of understanding but of *making* the world. And echoing Derrida’s points above, the power of these classificatory systems derives from their claim to innocence; from the notion that metadata are neutral descriptors of data.

Second, increasingly researchers use metadata as a portal into, and way of navigating, a dataset. Metadata enable researchers to interpret variables without having to access the whole dataset. Such a practice is not encouraged by the archive. We can see in the bottom the warning: “The frequencies displayed are unweighted. All results need careful interpretation. Original data collectors, depositors and the data archive bear no responsibility for any results or interpretations arising from the secondary use of the data.” Nevertheless when researchers use this, rather commonplace, practice they are working with (re)constituted data: data that have already been (re)made through metadata preparation practices. Indeed, they are working with data that is being manipulated through multiple practices: including but not restricted to data collection, data curation, and data analysis. While these practices are taken for granted and normatively treated as neutral, we argue that they all help constitute data *and* metadata. Neither data nor metadata are ready-made packets of meaningful information. Both are made through historically and culturally-specific practices that embed and enact particular conceptual assumptions that help constitute the nature and meaning of data and metadata.

³ see <http://www.data-archive.ac.uk/create-manage/consent-ethics/anonymisation?index=1> (Accessed 2015 May 31).

Data curation practices embody specific normative historically and culturally-contingent definitions of what is ‘meaningful’; they re-enact and re-inforce normative social science practice including typical scientific questions that might arise about a dataset and how analysis is usually performed. All this contributes to the definition of possible social scientific questions and methods and thus to a paradigmatic empirical social scientific field (Kuhn 1962).

7. Conclusions

Normative practices in archival and social science treat archived data, records and artefacts as representations of reality. We have sought to challenge this view through an examination of specific data curation practices in the social sciences. Archived data, we suggest, are made and remade through multiple practices including data collection, data curation and data analysis. In particular, we have suggested that

data curation, and other, practices *make* data through the conceptual commitments they embed and enact. Data curation (and other) practices are normatively treated as neutral and innocent, and their conceptual assumptions are inherited and taken as given and presuppositional. This renders invisible their constitutive role in the ontological formation of data, records and artefacts. Social scientific data curation practices (re)enact and (re)enforce social science practices that are already in place. Following Derrida, the power of archival and data curation practices lies in their power to bring together and materialise a conceptual configuration that is ‘at once visible and invisible’ (Derrida 1995:3) – that is both *specific* and naturalized. As others have argued (e.g. Somers 1996, 2008) social science has a role to play in opening up these conceptual configurations for empirical and theoretical debate and investigation. As Harris (2002:84) suggests, in failing to do so ‘we deny our audience the very space in which democracy thrives’.

8. APPENDIX 1

Metadata sheet form CESSDA

Yhteiskuntatieteellinen tietoaarkisto
Finnish Social Science Data Archive

DESCRIPTION TABULATION

Dataset: Election Funding of Finnish MPs 2007

Variable Q14 : [q14] The amount of contributions received from party organisations (in euros)

Literal Question
The amount of contributions received from party organisations (in euros)

Summary Statistics

Valid cases	271
Standard deviation	8086.337
Minimum	0.0
Maximum	57300.0
Mean	4601.3

Yhteystiedot FSD | FSD contacts | fsd@uta.fi

Copyright © 2001-2006 NSD - All Rights Reserved - Contact Nessstar Support

POWERED BY **nesstor**

9. REFERENCES

- [1] Andrews, M. 2008. Never the last word: Revisiting data. In *Doing narrative research*, M. Andrews, C. Squire, & M. Tamboukou, Eds., London, Sage, 205–222.
- [2] Barad, K. 2007. *Meeting the Universe Halfway*. Duke University Press.
- [3] Bornat, J. 2003. A second take: Revisiting interviews with a different purpose. *Oral History*, 31 (2003), 47–53. Checkland P., Holwell S. 1998. *Information, systems and information systems: making sense of the field*. John Wiley & Sons.
- [4] Cook, T. 2001. Archival science and postmodernism: new formulations for old concepts. *Archival Science* 1 (2001), 3-24.
- [5] Cook, T. 2013. Evidence, memory, identity, and community: four shifting archival paradigms. *Archival Science*, 13 (2013), 95–120.
- [6] Cook, T. and Schwartz, J. M. 2002. Archives, Records, and Power: From (Postmodern) Theory to (Archival) Performance. *Archival Science* 2, 3 (2002), 171-185.
- [7] Corti, L. 2011. The European landscape of qualitative social research archives. *Forum: Qualitative Social Research*, 12, Art. 11 (2011). Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/1746/3248>.
- [8] Derrida, J. 1995. *Archive Fever*. University of Chicago Press.
- [9] Duff, Wendy M., Harris, V. B. 2002. Stories and Names: Archival Description as Narrating Records and Constructing Meanings. *Archival Science*, 2 (2002), 263-285.
- [10] Duranti, L. 2001. The impact of digital technology on archival science. *Archival Science* 2001 1, 1, 39-55.
- [11] Foucault, M. 1972. *The Archaeology of Knowledge and The Discourse on Language*, Pantheon Books, New York.
- [12] Hammersley, M. 1997. Qualitative data archiving: Some reflections on its prospects and problems. *Sociology*, 31 (1997), 131–142.
- [13] Hammersley, M. 2010. Can we re-use qualitative data via secondary analysis? *Sociological Research*

- Online, 15, 5 (2010). Retrieved from <http://www.socresonline.org.uk/15/1/5.html>
- [14] Harris V. 2002. The archival sliver: Power, memory, and archives in South Africa. *Archival Science* 2, 1-2 (2002), 63-86.
- [15] Hedstrom, M. 2002. Archives, memory, and interfaces with the past. *Archival Science* 2, 1-2 (2002), 21-43.
- [16] Higgins, S. 2008. The DCC Curation Lifecycle Model. *The International Journal of Digital Curation*, 1, 3 (2008), 134-140.
- [17] Jenkinson, H. 1947. The English archivist: a new profession. In *Selected writings of Sir Hilary Jenkinson*. 1980. R. H. Ellis, P. Walne, Eds. Allan Sutton, Gloucester. Issued with new introduction in 2003 by Eastwood TM. Society of American Archivists, Chicago.
- [18] Kaplan, E. 2002. Many paths to partial truths: Archives, anthropology, and the power of representation. *Archival Science* 2, 3-4 (2002), 209-220.
- [19] Kuhn, T. 1962. *The Structure of Scientific Revolutions*. The University of Chicago Press, Chicago.
- [20] Law, J. 2004. *After method: Mess in social science research*. Routledge, Abingdon.
- [21] Mauthner, N. 2015. 'The past was never simply there to begin with and the future is not simply what will unfold': a posthumanist performative approach to qualitative longitudinal research. *International Journal of Social Research Methodology*, 18, 3 (2015), 321-336, DOI: 10.1080/13645579.2015.1022298.
- [22] Mauthner, N. S. Forthcoming-a. The Listening Guide method of narrative analysis: Towards a posthumanist performative (re)configuration. In *Feminist narrative research: Opportunities and challenge*. J. Woodiwiss, K. Smith, and K. Lockwood, Eds., Palgrave Macmillan, Basingstoke.
- [23] Mauthner, N. S. Forthcoming-b. Interview practices as diffraction apparatuses of bodily production. In *Mattering: Feminism, Science and Materialism*. V. Pitts-Taylor, Ed. New York University Press, New York.
- [24] Mauthner, N. S. and Parry, O. 2009. Qualitative data preservation and sharing in the social sciences: On whose philosophical terms? *Australian Journal of Social Issues*, 44 (2009), 289-305.
- [25] Mauthner, N. S., Parry, O. 2013. Open access digital data sharing: Principles, policies and practices. *Social Epistemology*, 27 (2013), 47-67.
- [26] Mauthner, N., Parry, O., and Backett-Milburn, K. 1998. The data are out there, or are they? Implications for Archiving and Revisiting Qualitative Data, *Sociology*, 32 (1998), 733-745.
- [27] Millerand, F., Ribes, D., Baker, K. S., and Bowker, G. C. 2012. Making an Issue out of a Standard: Storytelling Practices in a Scientific Community. *Science, Technology, & Human Values*, 38, 1 (2012), 7-43.
- [28] Moore, N. 2006. The contexts of context: Broadening perspectives in the (re)use of qualitative data. *Methodological Innovations Online*, 1 (2006), 21-32.
- [29] Moore, N. 2007. (Re)using qualitative data. *Sociological Research Online*, 12, 3 (2007). Retrieved from <http://www.socresonline.org.uk/12/3/1.html>.
- [30] Mottier, V. 2005. The Interpretive Turn: History, Memory, and Storage in Qualitative Research [21 paragraphs]. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [On-line Journal], 6,2, Art. 33 (2005). Available at: <http://www.qualitative-research.net/index.php/fqs/article/view/456> [Date of access: May 28, 2015].
- [31] Parry, O., Mauthner, N. 2004. Whose data are they anyway? Practical, Legal and Ethical Issues in Archiving Qualitative Research Data. *Sociology*, 38 (2004), 139-152.
- [32] Poole, A. H. 2014. How has your science data grown? Digital curation and the human factor: a critical literature review. *Archival Science* 15,2 (2014), 101-139, DOI 10.1007/s10502-014-9236-y.
- [33] Rasmussen, K. B., Blank, G. 2007. The data documentation initiative: a preservation standard for research. *Archival Science* 7 (2007), 55-71. DOI: 10.1007/s10502-006-9036-0
- [34] Sana, M., Weinreb, A. 2008. Insiders, Outsiders, and the Editing of Inconsistent Survey Data. *Sociological Methods & Research*, 36, 4 (May 2008), 515-541.
- [35] Savage, M. 2005. Revisiting classic qualitative studies. *Forum: Qualitative Social Research*, 6, 1, Art. 31. Retrieved from <http://www.qualitative-research.net/index.php/fqs/article/view/502> [Date of access: May 28, 2015].
- [36] Savage, M. 2010. *Identities and social change in Britain since 1940: The politics of method*. Oxford University Press, Oxford.
- [37] Schwartz, J. M and Cook, T. 2002. Archives, Records, and Power: The Making of Modern Memory. *Archival Science* 2 (2002), 1-19.
- [38] Somers, M. 1996. What is sociology after the historic turn? Knowledge cultures, narrativity, and historical epistemologies. In *The Historic turn in the human sciences*. T. J. McDonald, Ed., University of Michigan, Ann Arbor, MI, 53-90.
- [39] Somers, M. 2008. *Genealogies of citizenship: Markets, statelessness, and the right to have rights*. Cambridge University Press, Cambridge.
- [40] Stoler, A. L. 2002. Colonial archives and the arts of governance. *Archival Science*, 2, 1-2 (2002), 87-109.
- [41] Tamboukou, M. 2014. Archival research: Unravelling space/time/matter entanglements and fragments. *Qualitative Research*, 14 (2014), 617-633.
- [42] Thomson, R. 2014. Possession: Research practice in the shadow of the archive. In *The craft of knowledge: Experiences of living with data*. C. Smart, J. Hockey, and A. James, Eds., Palgrave Macmillan, Basingstoke, 39-55.
- [43] Trace, C. B. 2002. What is recorded is never simply 'what happened': Record keeping in modern organizational culture. *Archival Science* 2, 1-2 (2002), 137-159.
- [44] Whatley, P., Brown, C. 2009. Along and against the grain: the philosophy of the archive. *Archival Science* 9 (2009), 127-131.

