Article

S|S|S

Social Studies of Science
2019, Vol. 49(4) 503–530

DOI: 10.1177/0306312719846557
journals.sagepub.com/home/sss
SAGE

Impossible, unknowable, accountable: Dramas and dilemmas of data law

Alison Cool
Department of Anthropology, University of Colorado Boulder, USA

Abstract

On May 25, 2018, the European Union's General Data Protection Regulation (GDPR) came into force. EU citizens are granted more control over personal data while companies and organizations are charged with increased responsibility enshrined in broad principles like transparency and accountability. Given the scope of the regulation, which aims to harmonize data practices across 28 member states with different concerns about data collection, the GDPR has significant consequences for individuals in the EU and globally. While the GDPR is primarily intended to regulate tech companies, it also has important implications for data use in scientific research. Drawing on ethnographic fieldwork with researchers, lawyers and legal scholars in Sweden, I argue that the GDPR's flexible accountability principle effectively encourages researchers to reflect on their ethical responsibility but can also become a source of anxiety and produce unexpected results. Many researchers I spoke with expressed profound uncertainty about 'impossible' legal requirements for research data use. Despite the availability of legal texts and interpretations, I suggest we should take researchers' concerns about 'unknowable' data law seriously. Many researchers' sense of legal ambiguity led them to rethink their data practices and themselves as ethical subjects through an orientation to what they imagined as the 'real people behind the data', variously formulated as a Swedish population desiring data use for social benefit or a transnational public eager for research results. The intentions attributed to people, populations and publics – whom researchers only encountered in the abstract form of data – lent ethical weight to various and sometimes conflicting decisions about data security and sharing. Ultimately, researchers' anxieties about their inability to discern the desires of the 'real people' lent new appeal to solutions, however flawed, that promised to alleviate the ethical burden of personal data.

Keywords

accountability, big data, data law, ethics, GDPR, Sweden

Correspondence to:
Alison Cool, Department of Anthropology, University of Colorado Boulder, 1350 Pleasant Street, Hale Science 350 / 233 UCB, Boulder, CO 80309, USA.
Email: alison.cool@colorado.edu

On May 25, 2018, the European Union's General Data Protection Regulation (GDPR) came into force. Widely understood as 'the most influential piece of data protection legislation ever enacted' (Kuner et al., 2017), the GDPR grants individuals more control over their personal data while charging companies, institutions and organizations that process personal data with increased responsibility enshrined in principles like 'transparency' and 'accountability'. The vast scope of the GDPR, approved by the European Parliament in April 2016, has meant that almost every individual in the EU has been 'directly affected by the reform' (De Hert and Papakonstantinou, 2016: 180). With the GDPR, legislators are attempting to harmonize practices across 28 member states with a wide range of priorities and concerns about data.

Contemporary approaches to data protection are intimately connected to national histories of population data collection. As efforts to keep records of people who could be taxed or conscripted gradually evolved into national statistical institutions and population registers, different styles of enumeration emerged across Europe, such that '[e]very state, happy or unhappy, was statistical in its own way' (Hacking, 1990: 16). With the development of sophisticated bureaucracies for tracking populations came new forms of control. The precision of German, Polish and Dutch population data helped Nazis achieve deadly efficiency; France initiated far-reaching initiatives to collect population data under German occupation (Seltzer and Anderson, 2001). The Second World War has cast a long shadow on data practices in Europe. In France and Germany in particular, centralized national data collection and interlinkage through personal identification numbers are still viewed as problematic and suspicious (Poulain and Herm, 2013). In the Nordic countries, however, a strong historical relationship between the social welfare state, national population registries and data-driven policy has provided another lens through which extensive and centralized data collection appears if not beneficial, then at least not inherently negative.

In the 1970s and 1980s, different approaches to the politics of data collection in Europe were accommodated by varied national legislation (Bennett, 1992; Flaherty, 1992). By the early 1990s, there was an increasing sense that discrepancies between national data laws were creating obstacles to free trade and consumer protection.[1] The 1995 EU Data Protection Directive was intended to address this regulatory imbalance, but was transposed into national law in very different ways, contributing to inconsistency in policy and practice throughout the region (Bamberger and Mulligan, 2015). Unlike the 1995 Directive, the GDPR is directly and uniformly applicable throughout the EU – a strong statement that data protection has changed from a national prerogative to a European regulatory concern (De Hert and Papakonstantinou, 2016: 102).[2]

The window between the GDPR's approval in 2016 and its coming into force in 2018 allowed time for companies, institutions and organizations that process personal data in the EU or about EU residents to come to terms with the new requirements. This two-year period was charged with additional dramatic force by the steep fines for noncompliance – up to 20 million euros or four percent of annual turnover – that would be imposed when the GDPR came into effect in 2018. As Nicklas Lundblad, Google's Vice President of public policy and government relations for Europe, the Middle East and Africa, wryly observed at a seminar I attended at Stockholm's Centre for Business and Policy Studies in February 2017, the GDPR's four percent fine had a way of 'focusing the mind'. While

the GDPR is certainly and perhaps primarily intended to make tech platforms like Facebook and Google focus carefully on how they process personal data, the regulation applies to all entities that process personal data, and for this reason has also caught the attention of many people working at European scientific institutes, universities and research infrastructures.[3] Academic researchers share some concerns with corporate actors, including the common view that the GDPR's requirements are 'cumbersome and time-consuming' (Tikkinen-Piri et al., 2018: 135), but also experience the different challenge of aligning the new standards with existing practices and understandings of research ethics (Litton, 2017). In either case, GDPR compliance is further complicated by the need to reconcile national traditions of data collection with the new EU-wide legislation.

This article draws on research conducted in Sweden during the anticipatory period between the approval and implementation of the GDPR, and examines ethical challenges that emerged 'in action' as researchers measured their own data practices against what they understood as the 'in the books' requirements of data law.[4] Between January and April of 2017, I conducted ethnographic fieldwork in Stockholm, Uppsala and Göteborg, interviewing and interacting with researchers who work with population data, IT lawyers and legal scholars, and other experts in the areas of data regulation and data ethics.[5] The empirical data presented in this article focuses on these researchers' specific and varied experiences of managing legal and ethical uncertainty in Sweden, which I suggest can be productively contrasted with the broad ethical principles and universalizing tendencies of transnational data protection regulation such as the GDPR.

During my fieldwork, many of the researchers I talked to described a widespread sense of uncertainty about data law. Given the EU's ongoing and vexed efforts to harmonize data practices and the rapidly approaching implementation of the GDPR, a certain lack of clarity was probably to be expected.[6] Even so, I was struck by the profound doubt some researchers expressed as they talked about legal requirements they believed were 'impossible' to follow. A few researchers even suggested that when it came to data protection, *nobody* knew what the law required. On the surface, these claims are easy to dismiss. After all, anyone with an internet connection can read the regulation: the full text of the GDPR is available online in 24 languages, including Swedish, both as PDF and in html (European Union, 2016). As part of this project, I also spoke with lawyers and legal scholars in Sweden who were extremely knowledgeable about data protection law in general and the GDPR in particular. For them, data law was complex, perhaps flawed, but certainly not unknowable.

Nevertheless, I argue that we should take these researchers' concerns about data law and ethics seriously. As I will suggest, their sense of the ambiguity of the law in general and accountability in particular was a starting point for 'ethics work' focused on facilitating data use and exchange in particular ways (Hoeyer et al., 2017). These ethical processes, guided by researchers' uncertain anticipation of legal and technological changes to come, prompted reflection on the values that grounded their decisions about data security and data sharing and that were materialized in the datasets and research infrastructures they designed. In this sense, the 'unknowable' law acted as a 'values lever' (Shilton, 2013) that effectively produced the context-specific ethical responses that the GDPR's accountability principle was intended to achieve.

In practice, however, researchers often reformulated accountability as an ethical responsibility posed as a pragmatic question: To whom were they accountable? This view of accountability cast researchers as actors and data subjects as those who were acted upon, suggesting that researchers were responsible for and accountable to people they only encountered in the highly mediated form of data. Searching for someone to whom they could hold themselves accountable, researchers developed ethical relationships with what they imagined as the 'real people behind the data' – the persons, publics, or populations they thought that their research data might represent. As researchers envisioned the 'reality' of their data in different ways, they attributed a variety of person-like qualities and social values to the data. In this way, data was endowed with regional or national identities (representing, for example, a Swedish population or a European public) and social and economic preferences (rational, collectivist and pro-science). Researchers could then align their data practices with the interests they ascribed to the particular 'real people' they envisioned, fashioning themselves as ethical subjects through this improvised form of accountability (Shaw and Armin, 2011). However, researchers often felt unsure about the legality of their decisions and questioned the adequacy of data protection regulation that seemed to them to rely on the informal ethical guesswork of unqualified individuals. Seeking to relieve themselves of what they saw as the undue burden of ethics work, some embraced anonymization as a preferred solution to the dilemma of personal data – despite their knowledge that the promise of available anonymization techniques was often 'broken' by technical limitations (Ohm, 2009).

What is the law?

'We often wonder, like, what does the law say about this? Nobody knows', Hugo, a researcher and coordinator for Sweden's national bioinformatics platform, explained. It was early February 2017 and Hugo was telling me about how he determined appropriate security measures for a system that stored sensitive genomic and proteomic data. A few days earlier, Emil, another researcher who also worked with the national bioinformatics platform, made a similar observation: 'I don't think the universities or any company or organization does a very good job of spreading knowledge of the importance and – actually, what is the law? And it might change with the new regulation [the GDPR]. Who knows?' In April, when I asked Torbjörn, a lecturer in bioinformatics and head of a research computing storage facility, about working with sensitive data, he told me, 'As I see it, the legal rules are very deliberately vague.' Reflecting on a previous project, he explained: 'There are no checklists. So how should I build a system that can handle sensitive data? The guidelines say, more or less, that *adequate* safety should be in place. Riiiight – what does *that* mean? And there were no clear answers in that sense. I found that very frustrating.' With the specter of significant changes on the horizon with the GDPR, Hugo, Emil and Torbjörn's uncertainty was understandable. But why, when talking about data law, did they invoke not just uncertainty, but unknowability?

Hugo, Emil and Torbjörn seemed to refer at once to the GDPR as a new and anxiety-provoking source of legal uncertainty and to 'the law' in a wider, metaphorical sense. Classic scholarship in legal anthropology has analyzed how the law, in this second sense, figures into ordinary social life, as in common formulations in which 'the law' is used as

a shorthand for 'a very complex aggregation of principles, norms, ideas, rules, practices and the agencies of legislation, administration, adjudication and enforcement, backed by political power and legitimacy' (Moore, 1973: 719). In such formulations, the law is described as both abstract and agentive – divorced from everyday life, yet able to act upon it. The relationship between law and society – imagined as separate, interacting systems – was a key problem for legal anthropologists. Studies revealed the intersections of legal and social systems as sources of productive tension, and attempts to resolve mismatches between changing social practices and fixed legal codes were taken up as the 'fertile dilemma of law' (Bohannan, 1965: 37). Seen in this light, the ambiguity of legal concepts could be a resource for managing contradiction while maintaining the legitimacy of the law. As Gluckman (1973: 326) argued, '[t]he 'certainty' of law resides in the 'uncertainty' of its basic concepts.'

While data protection is not the only domain in which 'law exists on paper, but not in practice' (Cloatre and Pickersgill, 2014: 440), what seems noteworthy here is the felt presence of 'the law' as a desired but missing arbiter of research practice. Even in lamenting the failure of the law to provide wanted guidance, Hugo, Emil and Torbjörn assigned it an unquestioned force, an agency notable in its absence. For them, as with many other researchers implicit faith in the certainty of the law in a broader sense seemed to coexist with a strong sense of uncertainty about data law. As Clas, a molecular biologist in charge of coordinating research data across different national platforms, put it, 'we need to have the rules, and how the laws are set up, and the government authorities for inspecting and enforcing the rules – [it] should be done in a way that's so transparent for the researchers that there's no doubt. It shouldn't be the case that researchers need to sit in their own little rooms and think I have no idea what's going to happen if I do that.' In his view, uncertainty about appropriate data security therefore indicated a problem with the law. As Clas interpreted the situation, the law failed, in this instance, to do what was necessary and could otherwise be expected.

But is data law really unknowable? There are other ways of interpreting researchers' statements about uncertainty. Erika, an associate professor in public law, often interacted with researchers. She attributed their legal uncertainty not to a failure of the law, but to researchers' incorrect interpretations and tendencies to circulate misinformation:

> I have been working with [researchers] in different multidisciplinary projects [where there] were so many misunderstandings concerning law, and horrible situations where I met professors from Karolinska, or Kungliga Tekniska Högskolan and other places, where they said, 'and the law says this!' [pounds fist on table] 'and it's horrible!' And I was like, 'Wellllllll, no, it doesn't.' And they were like, 'Yes it does!' And I said: 'Where have you seen that? Who has said that to you?' And they were like: 'What? Isn't it true?' They were talking to each other and deciding what the law said. And they were completely wrong.

Erika described going out of her way to meet with researchers to explain legal requirements for data use, which she saw as her job as an employee of the university and as part of being a good colleague. However, she believed that researchers reached out to one another rather than coming to her or other legal experts for answers, and the result was a proliferation of mistakes and misinterpretations. The law, for Erika, was knowable. The

problem, as she saw it, was that researchers lacked knowledge of the law and looked for it in the wrong places.

Other legal scholars saw the law as complex or even flawed in ways that resonated somewhat with researchers' descriptions. Lars-Göran, a professor of law and information technology, told me that the GDPR 'lacked rationality', but that he saw this as inevitable when working with text, which was always subject to multiple interpretations. He understood legal ambiguity to be especially prevalent in so-called technology-neutral legislation like the GDPR, which attempts to anticipate rapid transformations in the field it seeks to regulate. 'It has to be vague,' he explained, 'or else you would have to change it all the time'. Similarly, Margareta, a lawyer and privacy researcher, pointed out that, 'the idea behind drafting [technology-neutral legislation] is that it will withstand the test of time. You want to have broad principles, you don't want to refer specifically to technologies because technology is always moving.' She then used the example of a data transfer to explain how broad principles and conceptual murkiness about specific technologies and practices can subvert the possibility of straightforward interpretation of the GDPR:

> If you think about a transfer, I think data is like – what does that even mean? You push a button and data goes from A to B? Data is flowing in a constant state, it's everywhere. I don't know that it's like you push this button and it moves from Sweden to New York. I don't think it's as easy to pinpoint what exactly a data transfer is. So I do think that there are these assumptions about how technology is that don't exist anymore. But then I will say that it's very hard, like what are you supposed to do from a legislative perspective?

For Lars-Göran and Margareta, the ambiguity of the law – which researchers saw as its unknowability – was part of its strength and flexibility. If the law was vague or difficult to interpret, to them this signaled not the failure of the law, but the instability and unruliness of technology. Where researchers saw the law as profoundly incomprehensible, lawyers and legal scholars saw technology as the underlying source of uncertainty. The resulting questions, however, were the same: What does this mean? What are you supposed to do?

Data made for use

To understand why the GDPR might appear unknowable to researchers or ambiguous to legal scholars it is helpful to consider the history of population data collection and research in Sweden. The Nordic countries, including Sweden, have long traditions of recordkeeping. From unusually well-kept church parish records and detailed population statistics (Sköld, 2004) to today's internationally renowned population registries and biobanks (Mattsson, 2016), Sweden has long been known as a 'country that kept track of its population' (Kälvemark, 1977). If the 'counting of people' (Hutchinson, 1959: 82) and 'the health and well-being of the entire population' (Sundin et al., 2005) have long been important political objectives in Sweden, the particularly data-driven characteristics of contemporary Swedish life were also made possible by the mid-twentieth-century development of a strong social welfare state and the introduction of the personal identification number (*personnumret*) in 1947. Sweden's population registries, easily

cross-linked due to the extremely widespread use of the personal identification number, are similar to those found throughout the Nordic countries; the unique depth and detail of the region's population registers and the ubiquitous use of personal identification numbers has resulted in a distinctively data-driven Scandinavian style of epidemiological research (Bauer, 2014). The Nordic countries are well-known for the unusual extent of their population data (Hoeyer, this volume); Sweden, by the simple virtue of having the largest population of the Nordic countries, might have claims to having the most data.

Swedish population data has long informed research and policy, and national data collection has often been understood as beneficial for both the citizen and society. A widely shared idea of a reciprocal relationship between citizens, science and a benevolent Swedish welfare state (Rothstein, 2006) has lent legitimacy to extensive national databases. While outside observers have described Sweden as 'model surveillance society' (Flaherty, 1992: 4), Swedish descriptions have been more likely to emphasize that 'Sweden has long been an information society' (Karlström, 1986: 108–9). In Sweden, the state's collection and use of population data is often described as balanced by the public access principle (*offenlighetsprincipen*), citizens' far-reaching constitutional right to access government data (Anderson, 1973; Lundvik, 1983; Österdahl, 1998; Sandell et al., 2003). Rather than a source of potential danger, Swedish population data has figured in national and scientific discourses as a kind of 'object made for use', at once a product of expert engineering and a vehicle for the improvement of the everyday life of Swedish citizens (Asdal and Gradmann, 2014: 180; cf. Mack, 2017; Murphy, 2013: 118). The researchers I spoke with often invoked these discourses of Swedish recordkeeping and research in the service of citizens.

National representations and narratives – as STS work on biobanking suggests – can serve as productive resources for constructing populations (Busby and Martin, 2006; Gottweis et al., 2011; Hinterberger, 2012; Mitchell and Waldby, 2010). As nationalities are imagined in particular ways, samples and data are taken to be imbued with the same national characteristics, which can then be referred back to as a source of legitimacy by the experts and institutions who process the data (Burton, 2018; Tarkkala and Tupasela, 2018; Tupasela et al., 2015). Resonant with the findings of the scholars noted above, researchers' constructions of the Swedish population's desire for research were also attributed to Swedish data as data 'made for use'. By extension, data-driven research became a legitimate and desirable use for Swedish data.

If Sweden's extensive national data collections have proven advantageous for researchers, regulating the vast array of government and scientific datasets and their varied uses has been a longstanding national problem. In 1973, Sweden passed the world's first national data protection statute.[7] Sweden joined the European Union in 1995, and the Personal Data Act (PuL) and Personal Data Ordinance (PuF) passed in 1998 implemented the 1995 EU Data Protection Directive on a national level. However, questions were raised from the outset about the compatibility of Swedish law with the EU Directive and how restrictions on data access in the new Personal Data Act might impinge on Swedish citizens' rights to access government data (Steele, 2002: 25–28).

In addition to the Personal Data Act, Sweden has passed many other acts pertaining to data protection in specific administrative sectors and databases.[8] By the late 2000s, scholars and other observers increasingly expressed concerns about the complexity

and vagueness of Swedish data laws (Bygrave, 2010: 9) and the lack of clarity around the hierarchy and relationships between the different laws (Öman, 2010). Uncertainty about research use of data in Sweden came to a head in 2011, when the Data Inspection Authority found that LifeGene, a biobank and research infrastructure comprising Sweden's largest biomedical project, was in violation of the law. Data Inspection's decision was considered controversial, and initiated an acrimonious period of debate until a temporary law was put in place in 2013 for the express purpose of allowing LifeGene's work to continue (Cool, 2016). Despite the seeming resolution of the LifeGene debate, some of the researchers I spoke with in 2017 cited this case as indicative of ongoing problems in Sweden with interpreting and applying data law to research data practices in particular.

Given that one of the main arguments in favor of the GDPR was its potential to simplify and harmonize rules for data processing, it might seem to offer a welcome opportunity for clarification about the legal status of research data practices. Indeed, some researchers expressed hope that it would. However, the GDPR is by no means a simple document. The final text of the GDPR includes 173 recitals and 99 articles; by contrast, the 1995 Directive included 72 recitals and 34 articles. The length and complexity of the regulation reflects the drawn-out and contentious negotiations involved in its drafting and ultimate approval.[9] During the period of deliberation, one area of special concern for the Nordic countries was securing the legal status of their national population data registers and widespread use of personal identification numbers (SOU, 2017: 50, 78). Sweden in particular has been especially focused on enabling ongoing research use of its national population data (Melin, 2015; SOU, 2017: 50).

In the final regulation, data processing in scientific research, in comparison with other sectors, was 'given a relatively strong position' (Sjöberg, 2017: 61). Despite this, at least two of the requirements have significant implications for scientific research: the first is the principle of integrity and confidentiality, which mandates 'appropriate security of personal data' and the use of 'appropriate technical or organisational measures' (European Union, 2016: Article 5[1,f]), and the second is the principle of accountability (European Union, 2016: Article 5[2]), which 'refers to the respect of ethical standards as being part of the lawfulness of [data] processing in research' (Chassang, 2017: 12). The GDPR's accountability principle, and the way that it collapses the domains of the legal and ethical into one another, is emblematic of what the researchers I spoke with described as the unknowability of the law. As I will argue, the qualities that make accountability such a flexible and effective tool for provoking individual ethical responses are the same qualities that researchers foreground in their descriptions of the impossible burden of legal compliance.

Ambiguity and accountability

The concerns of the researchers I talked to in Sweden were forged at the intersection of national traditions of recordkeeping and data-driven research and the increasingly transnational scope of research governance and data protection law, as exemplified by the GDPR. In this way, they reflect tensions between long-standing Swedish data practices and EU-level regulation aimed at harmonizing regional data protection. However, the

ethical dilemmas of researchers in Sweden also speak to what scholars have identified as broader concerns and transformations in data-driven sciences in the era of big data (Benezra, 2016; Hogle, 2016; Rapp, 2015; Ruckenstein and Schüll, 2017). As data is continuously generated, aggregated and analyzed, older anxieties about security, privacy and property have reawakened while new questions about whether and how to protect big data have emerged (Cakici, 2013; boyd and Crawford, 2012; Cate and Mayer-Schönberger, 2013; Gehrke, 2012; Krotoszynski, 2014; Zook et al., 2017). Efforts to regulate digital technologies across expanding spaces rely on broad, flexible principles that can encompass current technological realities and anticipate potential developments – all the while making room for cultural, linguistic and political differences. Further, strengthening data protection is often in tension with other regulatory imperatives to increase scientific data sharing and promote open data frameworks (Kaye, 2012). These legal tensions and regulatory complexities help to explain the 'ethical turn' in data protection policy (Raab, 2016) and research governance (Harvey and Salter, 2012). Broad ethical principles like transparency and accountability are favored solutions in research and data policy for their perceived ability to shore up the public legitimacy of science (Machado and Granja, 2018) while allowing for flexibility across time and space (Carter and Marchant, 2011; cf. Martin, 1994).

Accountability, however, has a longer history in data protection law and European politics. In EU policy discourse, accountability is often described as a unobjectionable tool for fostering legitimacy, but despite the concept's neutral veneer, debates about accountability are also fundamentally debates about what the future of Europe should be (Fisher, 2004). Within European data governance, calls for accountability often index debates about how to reconcile economic incentives to create a frictionless digital single market with rights-based arguments for restricting data flows to protect citizens' privacy and dignity (Jones, 2017).

Accountability was first introduced as principle for data protection in Europe in 1980 with the (non-binding) OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, which described the data controller as someone who 'should be accountable' for implementing measures in line with the principles it set forth (OECD, 1980; Varotto, 2015). Following this, accountability began to appear intermittently in international data guidelines and law, and in 1995 was referenced in the EU Data Protection Directive (Varotto, 2015: 79). Accountability was addressed in more detail in 2010, when the Article 29 Data Protection Working Party, an expert advisory group established under the 1995 Data Protection Directive, wrote that in order to take data protection 'from 'theory to practice', legal requirements needed to be translated 'into real data protection measures'. 'Accountability-based mechanisms' that required data controllers to 'implement appropriate and effective measures' in compliance with the law and 'demonstrate this on request', they argued, could be effective tools for accomplishing data protection in practice (Article 29 Data Protection Working Party, 2010: 3). Citing a need for 'flexibility' and 'scalability', the working party used 'open language' to outline the accountability principle (Article 29 Data Protection Working Party, 2010: 14). The working party pointed out that fulfillment of the accountability principle was not coterminous with legal compliance, and acknowledged that the linguistic and legal ambiguity of the principle might produce uncertainty among data controllers.

The GDPR's accountability principle is defined in Article 5 as the data controller's responsibility for and ability to demonstrate compliance with the data processing principles of lawfulness, fairness and transparency; purpose limitation, data minimization; accuracy; storage limitation; and integrity and confidentiality. The GDPR also demonstrates a commitment to flexibility with the use of open-ended language, such as 'every reasonable step' and 'appropriate security', in describing the principles. Giovanni Buttarelli, the European Data Protection Supervisor, wrote in a 2016 editorial that accountability was a central concept in the GDPR. Noting that the English word accountability 'derived ultimately from the Latin *putare* meaning "to reckon", he suggests that, '[t]he concept [of accountability has been incorporated into data protection because the need for such 'reckoning' stretches across all sectors that handle personal information, including government, private, academia, commercial, and not for profit' (Buttarelli, 2016: 78).

If accountability demands that controllers and regulators 'reckon' with data protection, such reckoning – like the accountability-based mechanisms described by the Article 29 Data Protection Working Party – is assigned to an ethical realm that overlaps but is not coextensive with the legal realm. As Buttarelli (2016: 78) argues: 'Being accountable for data processing is not a substitute for compliance with the applicable legal obligations. It should be understood as an ethical responsibility.' Accountability, described here as an ethical responsibility, requires interpreting open-ended concepts – what does it mean to act 'reasonably' or 'appropriately'? – and being able to demonstrate to others – who may or may not share your interpretation of reasonable or appropriate action – that you have done so. The uncertain reckoning that underlies accountability as an ethical responsibility produces new uncertainties when accountability is also a legal principle, making it difficult to determine in advance – before an ultimate legal reckoning – what is or is not permitted.

In this sense, as scholarship on audit cultures suggests, when accountability principles are written into law, researchers are transformed into 'accountable persons' (Miller and Leary, 1987), expected to internalize legal expectations and externalize virtue through measurable performances of ethics (Poovey, 1998; Shore, 2008; Shore et al., 2015; Strathern, 2000b). Yet, as ethnographic work on ethics review procedures has shown, practice on the ground does not always map on to the standardizing logic of accountability as a principle (Douglas-Jones, 2015; Hedgecoe, 2012; Lederman, 2007). As regimes of accountability expand into the realm of data practices, STS scholars have described how researchers are made accountable and account for data in different ways. Accountability can emerge as a pragmatic response to the collaborative nature of data-intensive research (Hoeppe, 2014); at other times, external demands for accountability can be productive of novel ethical sensibilities, as in the emergence of practices that distributed accountability across a network of team members, technical systems and ethics board members in an EU project to design an ethical algorithmic system (Neyland, 2016). In this case, as in other efforts to produce algorithmic accountability, calls to account are also ways to contest and change practices (Seaver, 2017). In genomic research and biobanking, calls for accountability in principle can be transformed into 'performative accountability' (Arribas-Ayllon, 2010) in practice, as researchers respond by seeking out 'foolproof' standards to prevent 'ethical mishaps' (Lassiter et al., 2016: 74).

Other STS work, arguing for the inseparability of ethical and technical concerns, makes a normative case for particular forms of accountability, as in suggestions that '*all* individuals involved [in data science] need to take some responsibility for potential implications' (Leonelli, 2016: 6) or that 'those in charge of emerging technologies are democratically obliged to take concerns into account, and answer to concrete objections' (Leese, 2017: 14). These normative perspectives are quite similar to the way that the accountability principle is framed in the GDPR, which also invokes the need for those processing data to bear the weight of present and future ethical concerns. Here, I would like to suggest that accountability in data processing is necessary and important, but also take a critical perspective on how demands for accountability affect those who are called to account.

Improvising accountability

What does accountability mean for researchers in Sweden? Accounting procedures, when applied to ethical conduct, emphasize relations between individuals imagined as actors and others who are acted upon. Ethics, therefore, 're-describes accountability as a matter of responsibility towards those who will be affected by the outcome of certain actions' (Strathern, 2000a: 292). For this reason, accountability, when mapped onto individual behavior via ethics, raises the question of to whom one should be accountable. This question can be difficult to answer, not least because of the polyvalence of accountability. The various meanings of the term are significant because 'different versions of accountability correspond to differing underlying assumptions about the nature of the audiences that are implicated' (Woolgar and Neyland, 2013: 37). In other words, each version of accountability suggests different answers to the ethical question of to whom one should be accountable. While there are already many versions of accountability in English usage, further complications are introduced when the concept travels beyond its original linguistic and legal context. For example, consider the question of what 'accountability' means in Swedish.

The native Swedish speakers among the researchers, lawyers and legal scholars I interviewed and interacted with, when speaking English, did not use the English words 'accountability' or 'accountable' to describe their thoughts or actions. When speaking Swedish, their preferred terms seemed to be *ansvar* or *ansvarig*, which are also used in the Swedish translation of the GDPR in the places where accountability is used in the English version. The accountability principle of the English version of the GDPR thus appears in the Swedish translation as *ansvarsprincipen*. There is no close equivalent in Swedish for 'accountability' (Lindkvist and Llewellyn, 2003). *Ansvar* is more commonly translated to English as responsibility or liability, and when speaking English, the native Swedish speakers most often used the term responsibility when describing data protection.[10]

The difficulty of translating accountability into Swedish is not unique (Veselý, 2013). The Article 29 Data Protection Working Group discussed the problem of using a concept from 'the Anglo-Saxon world' that 'due mainly to differences in the legal systems' had no equivalent in most European languages (Article 29 Data Protection Working Party, 2010: 7–8). While the working group suggested that the challenge of translating

accountability could lead to different interpretations among member states, potentially undermining harmonization efforts, they argued that the general meaning of the concept, 'showing how responsibility is exercised and making this verifiable', was clear enough (Article 29 Data Protection Working Party, 2010: 8). Accountability, they concluded, might simply '[refer] to the 'implementation of data protection principles' (Article 29 Data Protection Working Party, 2010: 8). In the end, accountability would seem to consist of a tautology: the legal responsibility to follow the law. For many of the researchers I spoke with in Sweden, the reasonable response to this legal qua ethical question of accountability could be summarized as: Well, yes, but how?

If the law failed to provide a clear answer to the 'how' question of legal compliance, researchers nevertheless discussed an obligation to act responsibly. While they did not use the term accountability, their descriptions of how they arrived at decisions about responsible or ethical action resonated with the kinds of accounting procedures and relations that have been analyzed under the rubric of accountability. For example, the researchers I spoke with often discussed questions about security and data protection as questions of relations with data subjects whom they imagined and described as the 'real' and 'alive' people 'behind the data'. Although for researchers the perspectives of the data subjects were almost always imagined rather than elicited, their efforts to envision the flesh-and-blood realness of personal data allowed them to take 'a crucial step out of the first-person point of view' (Keane, 2014: 451). In this way, researchers' senses of the ethical were constituted by thinking through how their actions might appear to or act upon the persons ostensibly referred to in personal data. If data infrastructures rely on and are constituted through the imaginative work of data experts who build and sustain them (Mackenzie, 2003; Nadim, 2016), for the researchers I interviewed, this type of creative effort was also critical for connecting the abstract dilemmas of law and ethics to tangible questions concerning everyday data practices.

As researchers worked out for themselves what data subjects might desire, they relied on improvisational accounting-like procedures. However, researchers described the provisional and informal character of their everyday 'ethics work' – the practical and often emotional labor of managing data flows (Hoeyer et al., 2017) – as an ongoing source of anxiety. They asserted that the law should, but failed to, outline clear security procedures and codes of conduct that they could measure their actions against. While some expressed hope that greater legal clarity might come after the implementation of the GDPR in 2018, they nonetheless found themselves, for the present, making decisions about data protection based on what they described as 'guessing' and 'balancing'.

Guessing and balancing

Because researchers situated their state of uncertainty within an informational abyss – after all, 'nobody knows what the law says' – they described feeling compelled to make their own interpretations of data laws. Many perceived this as an uncomfortable and difficult situation. Researchers thought that their interpretations formed a shaky framework for defining acceptable data practices, so they described actions based on these interpretations not as decisions, but as guesses. In their view, guessing was problematic, but

necessary. For example, Emil talked about the problem of formulating data security for the national bioinformatics infrastructure based on best guesses:

> It's a bit of a problem in that I think the law and the practice have not hashed out how things should work. How should we do things? And so we're on our end sort of guessing and trying to do the best we can, but we don't have a grasp of all the implications for this huge amount of data, how best to use it and there are international associations talking about and discussing these things, but still. But then the law – you don't know anything about the law until something has gone wrong and it's been in the courts. No one will tell you, well this is okay, this is not.

Similarly, Hugo described the process of determining appropriate security procedures for genomic data as 'kind of our best guess' and Torbjörn told me that he and his colleagues were 'not the only ones who are uncertain about [processing sensitive data]. And I mean the GDPR will come into effect, but it will be taken to court, probably, once or twice before it sort of settles on these things. Yeah, it's difficult, and, well, we try to do our best, I think.'

Best guesses – as researchers readily acknowledged – were not fully informed decisions or flawlessly reasoned judgments; rather, they were provisional solutions that allowed data-driven research to move forward. In their accounts, best guesses emerged from 'balancing', a pragmatic procedure in which researchers considered potential security and privacy risks in relation to the possible benefits of research. In this sense, 'balancing' can be seen as an informal accounting procedure that borrows some of the logic, if not the rigor, of a cost-benefit analysis. As researchers improvised, they also configured relations of accountability as they took into consideration the populations who might benefit from data-driven research and the persons to whom they imagined themselves responsible. Considerations about data security may not be an obvious site for articulating concerns about persons or imagining potential benefits for populations; yet, as researchers described it, these were indeed the objects of their processes of balancing and guessing.

Balancing is a process that takes the law, as researchers understand it, into account. However, balancing is not primarily about following the law, but instead is a form of 'ethical self-fashioning' (Shaw and Armin, 2011), a way of transforming oneself into a responsible, if not necessarily law-abiding, subject. For many researchers, responsibility emerges in relation to the research subjects from whom data originates. From a security perspective, the balance to be achieved is to make data accessible and usable for research while simultaneously protecting those whose bodies relinquished it from possible harms. As Hugo put it, 'Of course there is a tension there. On the one hand you want to make as much data as possible available, I mean that's always my instinct, but on the other hand, this is also personal information about actual subjects who are alive and living their lives.' The conflict here is that what Hugo sees as good for science – making data accessible – might be incompatible with adequately protecting the personal information of the 'actual subjects' to whom he says he feels responsible.

For many researchers, this tension cannot be resolved by guidelines-based data policy like the GDPR because it does not clearly define appropriate levels of protection for personal data. This is in many ways the intention behind the law – given its geographic

scope and the range of scenarios it encompasses, it makes sense to leave room for interpretation by establishing that the 'meaning of "appropriate" depends on the context' (Wolters, 2017: 172). Likewise, the accountability principle shifts responsibility to those who work with data, but at the same time 'allows for more flexible use of data' (Thomas, 2014: 139). Nevertheless, researchers saw this as a dilemma because if security was too lax, personal information about research subjects might be at risk, but if security procedures were too restrictive, research progress could be slowed or stopped. Legal accountability was understood as focused on risks of the first type, but for researchers, the second type of risk was also an accountability problem.

For example, Clas explained that 'most people are aware that you need special security, data security for [sensitive data], but there's also the balance in the worry that the rules are so complicated that by entirely following the law … research will not get done.' Clas felt some sense of legal responsibility, but perhaps more importantly, he explained, he felt responsible to the patients whose biological samples produced the genomic and proteomic data. 'You have a responsibility for the data not only to the government, but to the patients who actually gave their consent, your responsibility is to work really hard to make sure that you get the research out, that these are people who donated parts of their bodies for science.' From Clas's perspective, his responsibility to research subjects who gave their body parts 'for science' meant that he had an ethical duty to make the resulting data usable and accessible. This required balancing what Clas saw as the perceived desire of research subjects for their data to be used against the ostensibly byzantine demands of data law. For Clas, where the law seemed unclear, the ethical relationship between researchers and research subjects seemed more straightforward and more urgent.

Similarly, Jaime, a neuroscience researcher, talked about the research subjects to whom he felt accountable. As he explained: 'It's an ethical obligation, I think, to our research participants … to try to get the highest value out of the data set for which they have undertaken risks and harms.' The law, as Jaime understood it, was unnecessarily restrictive, due to what he saw as a tendency among legislators to assess 'risk from an unduly legalistic point of view. I think that harms research, and by harming research, it also harms people.' In contrast, Jaime described how balancing allowed him to determine that his ethical responsibility to research subjects would be best fulfilled by publishing his research data openly:

> We have some genotyping in a data set that we've put up now, and that's slightly – well, it's somewhat sensitive. It's not terribly sensitive, but it's somewhat sensitive, and we don't want the participants to find out about their genotype except if they ask us, because then they have the right to know, but we want to be able to tell them in such a way that we can manage any follow-up questions about medical risk that they might have. So in the end, after all this reasoning, we have determined that the risk to the participants is reasonable and is balanced by the benefit to science in general that is obtained by publishing the data openly.

Jaime concluded that research subjects' decisions to participate in research – despite any risks that may have been involved – was indicative of their desire for their data to be made available in the way that would have the most value for researchers.

Opening accountability

What Jaime and other researchers described as the value and benefits of publishing and sharing research data – which allowed them to frame it as an ethical practice – resonates with the Open Science movement, a diverse group of actors and organizations that encourage research and publishing practices that facilitate sharing data, results and research products as widely and accessibly as possible. In recent years, the European Commission has increasingly supported research and innovation policies based on ideas of open science and open innovation (European and Commission, 2015, 2016). The European Commission's embrace of open science and open innovation reveals anxieties about 'getting research results to market' (European Commission, 2016: 11) and the hope that data sharing will strengthen the EU economy and allow it to emerge as a 'stronger Global Actor' (European Commission, 2016: 5).

As in EU data protection policy, the EU open data framework incorporates accountability as a key principle, as in suggestions that: 'In being open, science will be fully accountable for its use of public resources' (European Commission, 2016: 52). Here, ethical notions of accountability are linked with an idea of researchers' fiduciary responsibilities as recipients of public resources. As with calls for accountability in research more broadly, studies of how researchers understand and implement open science principles of accountability have described a lack of clarity on the ground (Mauthner and Parry, 2013). Broad guidelines that urge openness in science without specifying what should be shared or how to do so 'generate considerable confusion and disagreement among researchers' (Levin et al., 2016: 130). Yet, for some of the researchers I interviewed, accountability in open science, as in data protection, was reframed as a relational question: To whom were they accountable? Here, the idea of the 'public' – a vaguely bounded entity that can nevertheless offer legitimacy to and benefit from research (Hinterberger, 2012; Pálsson and Prainsack, 2011) – provided a way for researchers to orient their ethical actions and reasoning (Venkat, 2017). As Emil put it, 'if it's publicly funded, the public has a right to it; they've paid for it.' Similarly, Jaime explained that 'they give us money, and they want to see us use the money wisely and not throw out the results or leave them to rot in a drawer somewhere.' Like the 'real people behind the data', the public is imagined as having rights and desires about research data that ethical researchers should try to discern. If real people were understood to have indicated their preference for data to be used by participating in research, the public was seen as akin to a group of shareholders who invested in research (by paying taxes) and therefore expected researchers to act as responsible stewards of their investment by minimizing waste (like data rotting in a drawer) and maximizing returns (by publishing data openly so it could enable the ongoing production of research results).

For researchers who understood sharing and publishing data as an ethical obligation to the public, uncertainty about data law could be an obstacle. As Emil explained, 'the biggest problem I think we have now is the sharing of data, or putting things in a repository – how is that done legally?' On the other hand, the combination of what looked like an ethical responsibility to share data and what seemed like the unknowability of the law could legitimize researchers' decisions to publish data sets. For example, Hugo described what he saw as 'a lot of confusion in the world of genomics' about publishing sequencing

data in public repositories like the Gene Expression Omnibus and the European Nucleotide Archive. 'A lot of these repositories where people are sending data, in many cases, they just submit, for instance, RNA sequences – they're from people, and I mean, you're *really* not supposed to do that. But it's like – no one knows how you should handle it, so they just keep on doing it.' Here, the seemingly unknowable law led researchers to conflicting formulations of ethical responsibility. Accountability, imagined in relation to a public eager to see the research results 'they paid for', suggests sharing and publishing as an ethical approach to data. In contrast, accountability to the people who provided samples might suggest either an ethical obligation to make data easily available to other researchers by publishing it openly, or that 'you're really not supposed to do that'. So what do the people and publics behind the data really want?

The real people behind the data

As these researchers' descriptions suggest, what they perceived as the 'unknowability' of the law ended up serving as an effective 'values lever' (Shilton, 2013), prompting them to reflect on what they considered valuable and worthy of protection. Amid a haze of legal uncertainty, the 'real' people 'behind the data' and the public 'who paid for it' seemed to offer a tangible focus for researchers' ethical concerns. Drawing on the long-standing national framework in which Swedish population data is 'made for use', accruing benefits for citizens by informing science and improving society, some of the researchers I spoke with came to see easing restrictions on data flows to facilitate its use in research as a way to enact ethical responsibility to the real people (assumed in this case to be Swedish) who provided samples or information. The EU Open Science framework, which highlights accountability to the public (assumed here to transcend national borders), helped other researchers to see depositing data in public repositories as ethical. However, the real persons or publics to whom researchers oriented their ethical actions were encountered only through their elusive representations as data and in the form of researchers' desires projected outward and reflected back to them. The relationship between the data at hand and the persons, populations or publics it is meant to represent and instantiate is not always as direct or obvious as it might seem.

For example, one of the projects that Hugo was working on was a large-scale cohort study focused on cardiovascular disease. This is how he described the process of data collection and sharing within the study:

> Usually the biological samples are taken by a nurse in Gothenburg, and she prepares the material that we need, and sends it to some facility, usually here in Stockholm, but some things are done in Umeå. Then the lab here, depending on which facility it is, they get the sample, and then they do whatever preparations they need to do, and put it into their machine, for instance, DNA sequencing or whatever it is, and then they deliver the data to us in some way. … If it's deemed to be sensitive, which is basically DNA or RNA sequence data, they deliver it from that facility into a secure system … You have to have both a password that you selected, and this code that you generate each time, and then you log in to that … system with the DNA and RNA data.

Hugo explained that he and his colleagues decided to use 'the strictest possible solution' for data security 'because it's individuals' data'. While Hugo described an ethical

responsibility to the 'actual' people who provided the samples, it is important to note that his interactions were not with those people, but with data that had gone through multiple displacements and translations before it arrived in the secure system where Hugo accessed it. The data, by virtue of having appeared in the system, indexed people's consent and enrollment in the study, but did it also convey the intentions that Hugo and other researchers might later attribute to the research subjects?

While Hugo worked with data collected specifically for the study, researchers who work with register data also generally operate at a substantial remove from the individuals who provide samples. Part of Clas's work was with a service platform for delivering genomic data. When I asked where the data came from, he told me that 'quite a lot of the samples come from Swedish national registers'. These samples were collected at different times and in various contexts from routine healthcare settings to research studies, where consent was understood and formulated in different ways. Clas pointed out that 'you can't go back to samples that were collected in the 70s and see that you have consent explicitly for whole genome sequencing'. Other procedures, like review by an ethical committee, can then be used to triangulate what these patients might once have understood or expected to happen to the biological materials they left behind. Balancing provided researchers with another ethical perspective.

My intention here is not to suggest that Hugo or Clas or other researchers I spoke with were incorrect or unethical in their assessments of patients' or research subjects' wishes to have their data used in research or made accessible to researchers, but rather to suggest that their access to the desires of the 'real' people 'behind the data' or the 'public' is not as straightforward as it might seem. Nevertheless, calls for accountability appear to encourage some researchers to summon up someone to whom they could be accountable. Given that their day-to-day work with data was unrevealing of the motives and relations that produced the data, researchers drew on available discursive frameworks (e.g., Swedish data made for use, the public interest) and their own capacities to imagine what the ever-elusive data subject would want. In almost all cases, the interests imputed to the imagined data subjects were closely aligned with researchers' own objectives such that the ethical course of action was to make data available for research.

These researchers proved adept at ethics work and skillful in ethical self-fashioning. They found productive ways to manage ambiguity and uncertainty, responding to laws they found unknowable by guessing, balancing and dealing with data in the best way they could. Reflecting on what data subjects might want, researchers identified their own values – openness, hard work, commitment to moving research forward – and recognized the impossibility of separating ethical questions from technical decisions. Still, it is difficult to ignore the strongly affective aspects of how many researchers talked about their work with data. Again and again, they described themselves using terms like 'frustrated', 'scared', 'unsure' and 'worried'. Researchers told me they felt surrounded by 'fear and inertia' in a situation they saw as if not 'impossible' or a 'quagmire', then certainly 'complicated', 'hard', 'problematic', 'difficult' and 'messy'. Searching for guidance and answers, they said that they found 'no one'. Far from the checklists and clear guidelines they wanted, what seemed like 'unknowable' and 'deliberately vague' laws held them accountable to mysterious people obscured 'behind' the data. What should we make of all this?

These researchers' accounts of unease about their data practices can be seen as a form of 'data anxiety', an experience of uncertainty among those who live and work with digital data about its safety, accessibility and management in the present and future (Pink et al., 2018). This anxiety, focused on 'the kinds of data that ordinary people store on their personal and work computers' is linked to 'an anticipatory temporality, in that saving data indicates that data needs to exist in an as-yet-unknown but imagined future' (Pink et al., 2018: 10–12). The researchers I talked with expressed similar concerns about the future, but their responses were most affectively charged when they spoke about data law and the ethical responsibilities it might entail. In this sense, their emotional responses suggested not only data anxiety, but what I would describe as accountability anxiety, an uncertainty about their ability to understand or adequately respond to the legal requirements and ethical obligations of personal data. Their anxiety, then, was not only about data as such, but also about data as a conduit to someone who could be helped or harmed. Researchers' anxiety stemmed in part from uncertainty about practical questions like where data should be stored, how to build a secure system to handle data, or who should be allowed to access data. These ordinary data dilemmas, however, acquired a dramatic cast when demands for accountability led researchers to see themselves as responsible for and accountable to people whom they only encountered in the highly mediated form of data. What might these people want or desire?

Researchers, seeing no other immediate solution, made guesses about the intentions of data subjects, but they wondered why – as ambiguous legal principles like accountability seemed to suggest – decisions about security and data sharing should be based on researchers' guesses. Torbjörn reflected on the consistency and efficiency of all of this guessing: 'How many systems do we have in Sweden handling sensitive data? I mean, it must be thousands – many, many thousands of them, but, yeah, I think they all make their own interpretations.' Emil, after telling me about guessing about security and sharing of sensitive data, expressed doubt about his ability to shoulder ethical responsibility for genetic data and, by proxy, all the Swedes who might be at risk if connections could be made between them and the data. He explained, 'even we [Emil and his colleagues] don't have a grasp of all the implications for this huge amount of data.' Perhaps, he suggested, a collective conversation about how to handle this would be a more adequate response than researchers' individual guesses: 'Society needs to have a discussion about the implications of all of it … and this discussion is not happening.'

Anonymization as antidote

Frustrated with interpreting the law and guessing what data subjects wanted, some researchers sought to cast off the ethical weight of personal data. Here, the law finally appeared to suggest a clear solution. As Margareta explained, 'the number one rule with data protection law is that it's only when you're dealing with personal data, so if your data set is not personal data, if you have anonymized it, then you're good, you're outside of it.' The GDPR defines anonymous information – to which its principles do not apply – as 'information which does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable.' (European Union, 2016: Article 119[5]). Anonymization, therefore,

would appear to release researchers from legal responsibility and the burdensome ethical demands of the people, publics and populations 'behind the data'.

Like many other researchers, Sam, a computational biologist, said that he concerned about the risks of working with personal data, but found that the ambiguity of data protection law left him in an uncertain position: 'Everyone recognizes there is a privacy issue, but what does it really mean?' For one of his projects, a large-scale collaborative study of pandemics, Sam was creating a 'synthetic population', a kind of statistical model of the population of the city of Gothenburg. Despite its precision and detail, the synthetic population, according to Sam, posed no risk to any real individual:

> It's not a real population, it's anonymized and so on, to take care of those things, but it is a statistically accurate representation of a real population, and that's what you really need in order to study this question of how an infection might spread if it's communicated to someone who comes to Gothenburg, and then that person goes to take the kids to school, the infection is going to start from that point and spread.

However, to create the model, Sam needed detailed data about the population of Gothenburg, from locations of schools and workplaces to composition of households and mobility patterns of residents. Statistics Sweden, the national statistical office, controlled access to some of the population data that he needed. While Sam asserted that the synthetic population wouldn't 'compromise the privacy of any individual at all', Statistics Sweden saw things differently:

> The safest route for [Statistics Sweden] is to say no. The moment there is the slightest suspicion that there may be any compromise, they are going to be very – overly conservative. Whereas for us, it's a process of trying to explain to them that this is what we are doing with it, and it doesn't compromise any privacy, and it's going to have these benefits. So it's a constant back and forth between us.

Sam, confident that anonymization of the data would prevent any possible harm to the residents of Gothenburg, found it tedious to have to account for potential risks and benefits in order to gain access to the data. As he saw it, there was no need for this kind of balancing because the creation of a synthetic population already solved the ethical problem of protecting the 'real' population. In contrast to other researchers who described feeling uncertain or anxious about data protection, Sam seemed sure that anonymization was effective and that there were no real people to be found behind the synthetic data.

Sam was not alone in seeing the appeal of anonymization. Other researchers, however, were optimistic that anonymization could minimize but not eliminate risk, as Sam seemed to suggest. The creation of synthetic data is an anonymization technique based in the growing field of statistical disclosure control, which focuses on finding computational solutions to the challenges of data protection. Research in this area often aims to anonymize data according to the standards of differential privacy, which promises a 'robust, meaningful, and mathematically rigorous definition of privacy' (Dwork and Roth, 2014: 3). Differential privacy can be achieved in multiple ways, but generally involves an algorithm that introduces some randomness or noise into a dataset. The aim, broadly speaking, is to transform the dataset such that someone analyzing it would not be

able to learn anything 'about an individual' (as it would not be possible to tell if an individual data point were signal or noise), but would still be able to learn 'useful information about a population', as the transformed 'differentially private' dataset, as a whole, would reveal the same patterns as the original, despite the changes to some of the data (Dwork and Roth, 2014: 5). Differential privacy can be used, among other things, to generate synthetic data. The data is considered 'synthetic' because it was generated algorithmically, in contrast to 'real' data that is based on a measurement. Given that many of these researchers perceived the 'reality' of data – its link to 'real people' – as the source of difficult ethical obligations and the basis for impossible legal requirements, synthetic data would appear to resolve researchers' ethical and legal dilemmas about data protection. In other words, the appeal of anonymization for researchers is that a data subject who cannot be identified is also a data subject to whom a researcher cannot be held accountable.

The strength of anonymization's appeal sometimes outweighs knowledge of its shortcomings. Re-identification research strongly suggests that data 'can be either useful or perfectly anonymous but never both' (Ohm, 2009: 1704). Synthetic data approaches, for example, can make identification of personal data far more difficult, but increases in data protection come in inverse proportion to the accuracy of the data and the validity of inferences made from that data (Reiter, 2002: 532). Furthermore, synthetic population models of pandemics, like other *in silico* models with medical applications, rely on clinical or medical datasets both for parameterization and validation (Carusi, 2014, 2016). And, even with synthetic data, there are still decisions to be made about what constitutes adequate protection – after all, '[t]here are no universal standards for acceptable levels of identification risk (Loong et al., 2013: 4152). Awareness of anonymization's 'broken promise', however, seems to coexist with persistent faith in its ability to resolve the intractable tensions of data protection (Ohm, 2009; Tempini and Leonelli, 2018).

Margareta, for instance, initially suggested that anonymization was the best way for researchers to respond to the GDPR. Yet, when I asked her if anonymization of personal data was really possible, she sighed and said: 'Well, yeah, I mean, I think we both know the answer is no.' Similarly, while Sam first described differential privacy as 'an ironclad guarantee that no personal information gets released', he later acknowledged that the theoretical guarantee of these techniques had not quite been settled in real-world applications: 'Whether it actually translates into practice involves the tuning of various parameters … there are various hidden constants here and there that in theory are not that important, but in practice they are.' Here, once again, what first appears certain veers back into uncertainty.

Conclusion

What is real here? Is anything clear, knowable, certain? Vague yet urgent calls for accountability have proliferated with efforts to harmonize data regulation on a transnational scale. As the GDPR and other data regulations aim to expand digital markets without sacrificing privacy or security, accountability principles seem to offer a flexible solution. The ambiguity and open-endedness of accountability allow the concept to travel easily and lightly, accommodating variation in local practice and leaving space for

new technological possibilities. However, the qualities that make accountability an adaptable, portable and arguably effective legal principle can evoke anxiety among those who are held to account. In itself, this anxiety is not inherently problematic. Indeed, anxiety is perhaps the most successful mechanism for transforming people into accountable subjects. Yet, as Strathern (2000b) has argued, the ethics of accountability can produce effects that are simultaneously 'obstructive, destructive even, *and* vitalizing' (p. 14).

In this sense, what the researchers I spoke with in Sweden took to be the 'unknowability' of the law became a starting point for their ethical engagement with the 'real people behind the data' and encouraged them to reflect on the possible consequences of their decisions about security and data sharing. Imagining the wants and needs of the people, publics or populations from whom the data originated, researchers fashioned themselves as ethical subjects by taking these desires into account. Ultimately, however, researchers could only guess at what the 'real' people might want. Some, like Emil, who called for a larger discussion of the implications of big data, wondered if the 'real' people even knew what they wanted. Best guesses, then, often produced more anxiety than certainty for researchers. Personal data – by virtue of its fraught connection to persons – seemed, to them, to demand an unachievable accountability, a legal qua ethical responsibility to everyone and no one. Anonymization, at least in theory, provided a perfect antidote to their anxieties of accountability because anonymized data, by definition, could not be connected to any person. Indeed, I would argue that the continuing appeal of anonymization for researchers – despite their knowledge that the theory of anonymization did not bear out in practice – is in its promise to relieve them of the burdens of accountability.

In Sweden and globally, researchers are increasingly expected to be accountable, but to whom? And how? What is left unspecified in flexible legal principles must be resolved on a case-by-case basis by researchers as they work to reconcile national traditions of research data use and population-making with the global rerouting of data and accompanying demands for standardization. In contrast to flexible, universalizing narratives of accountability, national stories of how researchers go about the daily work of data protection in light of such demands can give perspective on how population data is variously formulated as an object of ethical and legal responsibility. These stories are important because the particular forms of responsibility that researchers assume, the desires and characteristics they assign to population data in specific national contexts (as in Swedish data as data made for use), and the anxieties they experience along the way inform their decisions about what kind of data protection is possible or desirable (as in enthusiasm about anonymization as a solution to the ethically problematic 'reality' of personal data). In turn, flows of research data are directed in ways that the law necessitates but perhaps does not anticipate.

Acknowledgements

I would like to thank Klaus Hoeyer, Martyn Pickersgill and Susanne Bauer for their profoundly generous and insightful editorial guidance throughout multiple iterations of this text. I am also deeply indebted to Donna Goldstein, Carla Jones, and participants in the ethnography working group at CU Boulder for their many helpful comments and suggestions. Lauren Barrett's flawless research assistance was invaluable for this project. I am grateful to the three anonymous reviewers and to Sergio Sismondo for their thoughtful engagement with my writing. Lastly, many thanks to

Anna Dreber Almenberg and Anna Bohlin for supporting my research in countless ways, and to my research participants in Sweden for their trust and their time.

Notes

1. In the 1970s, France, Germany, Denmark and Sweden – nations where data collection has been particularly politically salient – implemented some of the earliest national data protection laws. By the early 1990s, Greece, Italy, Portugal and Spain had yet to put in place any data protection legislation at the national level, despite the issuing of OECD guidelines in 1980, a Council of Europe Convention in 1985 and a UN Resolution in 1989 – all of which called for the implementation and coordination of national data laws (Banisar and Davies, 1999; Newman, 2008; Wuermeling, 1996).
2. The choice of a regulation, rather than a directive, as the policy instrument for the GDPR is extremely significant. In EU policy, directives (like the 1995 Data Protection Directive) offer guidelines that are transposed into domestic legislation in each member state, allowing considerable flexibility at the national level. Regulations like the GDPR, in contrast to directives, are directly and uniformly applicable in all member states and binding in their entirety. As a regulation, the GDPR was intended to harmonize data protection in the EU and reduce the uncertainty and ambiguity that characterized data law and practice under the 1995 Directive (Wagner and Benecke, 2016; Wolters, 2017).
3. In practice, distinctions between science and industry – and the allocation of interests and values to each category – are not clear-cut and instead emerge and shift through the ongoing 'relational work' of actors (Lee, 2015). In this sense, scientific data processing and corporate data processing are not predefined, discrete categories. Nor can a stable set of conventions, logics and standards be attributed to either. Nevertheless, the GDPR outlines some exceptions for data processing in scientific research even as its implementation and broad ethical principles pose significant and related challenges for both industry and academia.
4. This study underwent IRB approval at the University of Colorado Boulder. The ethnographic fieldwork included 61 semi-structured interviews and participant-observation at seminars, conferences, meetings and events focused on data, technology and legislation. Interview subjects were initially selected by identifying researchers and professionals actively publishing or presenting data-driven work and/or publicly discussing data protection, data infrastructure, data policy, data ethics, open data and big data. Additional interview subjects were identified using snowball sampling. Interviews with researchers included questions about what kinds of data they worked with in their research, how they made decisions about data storage, security and sharing, and what challenges they experienced in their work with data. Interviews with ethicists, lawyers and legal scholars included questions about the history of data law and research ethics in Sweden, intersections between Swedish and EU law, and their interactions with researchers. Interviews were primarily conducted in English with some shorter exchanges in Swedish; participant-observation took place at events where both Swedish and English were spoken. I transcribed recordings of interviews and then coded them thematically, along with my typed fieldnotes, using NVivo.
5. About half of the interview subjects (26) were researchers, including graduate students, postdocs and faculty, all of whom held academic positions at universities or university-affiliated research centers or infrastructures. Researchers are identified according to the department they were affiliated with at the time of my fieldwork, which included departments of computer science, information technology, bioinformatics, computational biology, genomics, neuroscience, economics and climate science. In addition to the 26 researchers, I also interviewed 14 lawyers, legal scholars and ethicists. The remaining 21 interview subjects included

journalists, activists, policymakers and other professionals for whom data law or data ethics was central to their work.

6. Uncertainty about the effects of EU policy on everyday practice is common; scholars have described this phenomenon as the 'black hole' (Mastenbroek, 2005) or 'black box of EU law in action' (Versluis, 2007: 54).
7. The Data Act (*datalagen*), which also established the national Data Inspection Authority, was passed after a series of investigations about 'computers and privacy' following debates surrounding the 1970 census (Flaherty, 1986: 7–8).
8. Examples include the Personal Data Ordinance (*Personuppgiftsförordningen*), the Patient Data Act (*Patientdatalagen*) and the Law on Ethical Review of Research Involving Humans (*Lag om etikprövning av forskning som avser människor*), to name just a few.
9. Between 2009, when the process began, and 2016, when the GDPR was approved, the European Parliament received 4000 amendment proposals and lobbying reached 'unprecedented levels' (De Hert and Papakonstantinou, 2016: 181). The GDPR is the outcome of many compromises across fault lines both old and new – including the longstanding 'North/South divide' among member states; (McDonald, 2000: 116); tensions between European Commission's emphasis on economic and security issues and the European Parliament's focus on individual rights (Burri and Schär, 2016); and the sometimes-conflicting interests of industry, the public sector and academia.
10. Furthermore, accountability is not a traditional political value in Sweden. Unlike the British approach, where the government's accountability to voters has indexed democracy, in Sweden, democratic legitimacy has traditionally been anchored in the values of representativeness (through shared power) and consensus (through finding solutions that are acceptable to all, not just a majority) (Lewin, 1998: 203).

References

Anderson SV (1973) Public access to government files in Sweden. *The American Journal of Comparative Law* 21(3): 419–473.

Arribas-Ayllon M (2010) Beyond pessimism: The dialectic of promise and complexity in genomic research. *Genomics, Society and Policy* 6(2): 1–12.

Article 29 and Data Protection Working Party (2010) Opinion 3/2010 on the principle of accountability. Brussels. Available at: https://www.dataprotection.ro/servlet/ViewDocument?id=654 (accessed 26 November 2018).

Asdal K and Gradmann C (2014) Introduction: Science, technology, medicine – And the state: The science-state nexus in Scandinavia, 1850–1980. *Science in Context* 27(2): 177–186.

Bamberger KA and Mulligan DK (2015) *Privacy on the Ground: Driving Corporate Behavior in the United States and Europe*. Cambridge, MA: The MIT Press.

Banisar D and Davies S (1999) Global trends in privacy protection: An international survey of privacy, data protection, and surveillance laws and developments. *John Marshall Journal of Computer and Information Law* 18: 1–112.

Bauer S (2014) From administrative infrastructure to biomedical resource: Danish population registries, the "Scandinavian laboratory", and the "epidemiologist's dream". *Science in Context* 27(2): 187–213.

Benezra A (2016) Datafying microbes: Malnutrition at the intersection of genomics and global health. *BioSocieties* 11(3): 334–351.

Bennett CJ (1992) *Regulating Privacy: Data Protection and Public Policy in Europe and the United States*. Ithaca, NY; London: Cornell University Press.

Bohannan P (1965) The differing realms of the law. *American Anthropologist* 67(6): 33–42.

boyd d and Crawford K (2012) Critical questions for big data. *Information Communication & Society* 15(5): 662–679.

Burri M and Schär R (2016) The reform of the EU data protection framework: Outlining key changes and assessing their fitness for a data-driven economy. *Journal of Information Policy* 6: 479–511.

Burton EK (2018) Narrating ethnicity and diversity in Middle Eastern national genome projects. *Social Studies of Science* 48(5): 762–786.

Busby H and Martin P (2006) Biobanks, national identity and imagined communities: The case of UK biobank. *Science as Culture* 15(3): 237–251.

Buttarelli G (2016) The EU GDPR as a clarion call for a new global digital gold standard. *International Data Privacy Law* 6(2): 77–78.

Bygrave LA (2010) Privacy and data protection in an international perspective. *Scandinavian Studies in Law* 56: 165–200.

Cakici B (2013) *The Informed Gaze: On the Implications of ICT-Based Surveillance*. Doctoral Dissertation, Stockholm University. Available at: http://urn.kb.se/resolve?urn=urn:nbn:se:su:diva-92956

Carter RB and Marchant GE (2011) Principles-based regulation and emerging technology. In: Marchant G, Allenby B and Herkert J (eds) *The Growing Gap Between Emerging Technologies and Legal-Ethical Oversight*. Dordrecht: Springer, 157–166.

Carusi A (2014) Validation and variability: Dual challenges on the path from systems biology to systems medicine. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 48: 28–37.

Carusi A (2016) In silico medicine: Social, technological and symbolic mediation. *Humana-Mente Journal of Philosophical Studies* 30: 67–86.

Cate FH and Mayer-Schönberger V (2013) Notice and consent in a world of big data. *International Data Privacy Law* 3(2): 67–73.

Chassang G (2017) The impact of the EU General Data Protection Regulation on scientific research. *ecancermedicalscience* 11: 709.

Cloatre E and Pickersgill M (2014) International law, public health, and the meanings of pharmaceuticalization. *New Genetics and Society* 33(4): 434–449.

Cool A (2016) Detaching data from the state: Biobanking and building big data in Sweden. *BioSocieties* 11(3): 277–195.

De Hert P and Papakonstantinou V (2016) The new General Data Protection Regulation: Still a sound system for the protection of individuals? *Computer Law & Security Review* 32(2): 179–194.

Douglas-Jones R (2015) A 'good' ethical review: Audit and professionalism in research ethics. *Social Anthropology* 23(1): 53–67.

Dwork C and Roth A (2014) The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9(3–4): 211–407.

European Commission (2015) *Access to and Preservation of Scientific Information in Europe*. Brussels: European Commission.

European Commission (2016) *Open Innovation, Open Science, Open to the World – A Vision for Europe*. Luxembourg: European Commission.

European Union (2016) General Data Protection Regulation. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32016R0679 (accessed 25 November 2018).

Fisher E (2004) The European Union in the age of accountability. *Oxford Journal of Legal Studies* 24(3): 495–515.

Flaherty DH (1986) Governmental surveillance and bureaucratic accountability: Data protection agencies in Western societies. *Science, Technology, & Human Values* 11(1): 7–18.

Flaherty DH (1992) *Protecting Privacy in Surveillance Societies: The Federal Republic of Germany, Sweden, France, Canada, and the United States*. Chapel Hill, NC; London: The University of North Carolina Press.

Gehrke J (2012) Quo vadis, data privacy? *Annals of the New York Academy of Sciences* 1260(1): 45–54.

Gluckman M (1973) *The Judicial Process among the Barotse of Northern Rhodesia*. Manchester: Manchester University Press.

Gottweis H, Chen H and Starkbaum J (2011) Biobanks and the phantom public. *Human Genetics* 130(3): 433–440.

Hacking I (1990) *The Taming of Chance*. Cambridge: Cambridge University Press.

Harvey A and Salter B (2012) Anticipatory governance: Bioethical expertise for human/animal chimeras. *Science as Culture* 21(3): 291–313.

Hedgecoe AM (2012) Trust and regulatory organisations: The role of local knowledge and face-work in research ethics review. *Social Studies of Science* 42(5): 662–683.

Hinterberger A (2012) Publics and populations: The politics of ancestry and exchange in genome science. *Science as Culture* 21(4): 528–549.

Hoeppe G (2014) Working data together: The accountability and reflexivity of digital astronomical practice. *Social Studies of Science* 44(2): 243–270.

Hoeyer K, Tupasela A and Rasmussen MB (2017) Ethics policies and ethics work in cross-national genetic research and data sharing: Flows, nonflows, and overflows. *Science, Technology, & Human Values* 42(3): 381–404.

Hogle LF (2016) Data-intensive resourcing in healthcare. *BioSocieties* 11(3): 372–393.

Hutchinson EP (1959) Swedish population thought in the Eighteenth Century. *Population Studies* 13(1): 81–102.

Jones ML (2017) The right to a human in the loop: Political constructions of computer automation and personhood. *Social Studies of Science* 47(2): 216–239.

Kälvemark AS (1977) The country that kept track of its population: Methodological aspects of Swedish population records. *Scandinavian Journal of History* 2(1): 211–230.

Karlström G (1986) Information systems in local governments in Sweden. *Computers, Environment and Urban Systems* 11(3): 107–113.

Kaye J (2012) The tension between data sharing and the protection of privacy in genomics research. *Annual Review of Genomics and Human Genetics* 13: 415–431.

Keane W (2014) Freedom, reflexivity, and the sheer everydayness of ethics. *HAU: Journal of Ethnographic Theory* 4(1): 443–457.

Krotoszynski RJJ (2014) Reconciling privacy and speech in the era of big data: A comparative legal analysis. *William & Mary Law Review* 56: 1279.

Kuner C, Jerker D, Svantesson B, et al. (2017) The GDPR as a chance to break down borders. *International Data Privacy Law* 7(4): 231–232.

Lassiter D, Cadigan RJ, Haldeman KM, et al. (2016) Standardization as performative accountability in biobanking. *BioSocieties* 11(1): 67–81.

Lederman R (2007) Comparative "research": A modest proposal concerning the object of ethics regulation. *Polar: Political and Legal Anthropology Review* 30(2): 305–327.

Lee F (2015) Purity and interest: On relational work and epistemic value in the biomedical sciences. In: Dussauge I, Helgesson C-F and Lee F (eds) *Value Practices in the Life Sciences and Medicine*. Oxford: Oxford University Press, 207–223.

Leese M (2017) Holding the project accountable: Research governance, ethics, and democracy. *Science and Engineering Ethics* 23(6): 1597–1616.

Leonelli S (2016) Locating ethics in data science: Responsibility and accountability in global and distributed knowledge production systems. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* 374(2083): 20160122.

Levin N, Leonelli S, Weckowska D, et al. (2016) How do scientists define openness? Exploring the relationship between open science policies and research practice. *Bulletin of Science, Technology & Society* 36(2): 128–141.

Lewin L (1998) Majoritarian and consensus democracy: The Swedish experience. *Scandinavian Political Studies* 21(3): 195–206.

Lindkvist L and Llewellyn S (2003) Accountability, responsibility and organization. *Scandinavian Journal of Management* 19(2): 251–273.

Litton J-E (2017) We must urgently clarify data-sharing rules. *Nature News* 541(7638): 437.

Loong B, Zaslavsky AM, He Y, et al. (2013) Disclosure control using partially synthetic data for large-scale health surveys, with applications to CanCORS. *Statistics in Medicine* 32(24): 4139–4161.

Lundvik U (1983) The public's access to official documents in Sweden: The rules and their consequences. *Government Publications Review* 10(1): 3–9.

McDonald M (2000) Accountability, anthropology and the European Commission. In: Strathern M (ed.) *Audit Cultures: Anthropological Studies in Accountability, Ethics and the Academy*. London; New York: Routledge, 106–132.

Machado H and Granja R (2018) Ethics in transnational forensic DNA data exchange in the EU: Constructing boundaries and managing controversies. *Science as Culture* 27(2): 242–264.

Mack J (2017) *The Construction of Equality: Syriac Immigration and the Swedish City*. Minneapolis, MN: University of Minnesota Press.

Mackenzie A (2003) These things called systems: Collective imaginings and infrastructural software. *Social Studies of Science* 33(3): 365–387.

Martin E (1994) *Flexible Bodies: Tracking Immunity in American Culture from the Days of Polio to the Age of AIDS*. Boston, MA: Beacon Press.

Mastenbroek E (2005) EU compliance: Still a 'black hole'? *Journal of European Public Policy* 12(6): 1103–1120.

Mattsson T (2016) Quality registries in Sweden, healthcare improvements and elderly persons with cognitive impairments. *European Journal of Health Law* 23(5): 453–469.

Mauthner NS and Parry O (2013) Open access digital data sharing: Principles, policies and practices. *Social Epistemology* 27(1): 47–67.

Melin AS (2015) Sverige huvudaktör i uppgörelse. *Dagens Nyheter*, 16 June. Available at: https://www.dn.se/arkiv/nyheter/sverige-huvudaktor-i-uppgorelse/

Miller P and O'Leary T (1987) Accounting and the construction of the governable person. *Accounting, Organizations and Society* 12(3): 235–265.

Mitchell R and Waldby C (2010) National biobanks: Clinical labor, risk production, and the creation of biovalue. *Science, Technology, & Human Values* 35(3): 330–355.

Moore SF (1973) Law and social change: The semi-autonomous social field as an appropriate subject of study. *Law & Society Review* 7(4): 719–746.

Murphy KM (2013) A cultural geometry: Designing political things in Sweden. *American Ethnologist* 40(1): 118–131.

Nadim T (2016) Data labours: How the sequence databases GenBank and EMBL-Bank make data. *Science as Culture* 25(4): 496–519.

Newman AL (2008) Building transnational civil liberties: Transgovernmental entrepreneurs and the European Data Privacy Directive. *International Organization* 62(1): 103–130.

Neyland D (2016) Bearing account-able witness to the ethical algorithmic system. *Science, Technology & Human Values* 41(1): 50–76.

Ohm P (2009) Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review* 57: 1701–1778.

Öman S (2010) Trends in data protection law. *Scandinavian Studies in Law* 56: 210–205.

Organisation for Economic Co-operation and Development (OECD) (1980) *Guidelines on the Protection of Privacy and Transborder Flows of Personal Data*. 23 September, Paris, France. Available at: https://www.oecd.org/internet/ieconomy/oecdguidelinesontheprotectionofprivacyandtransborderflowsofpersonaldatabackground.htm

Österdahl I (1998) Openness v. secrecy: Public access to documents in Sweden and the European Union. *European Law Review* 23: 336–356.

Pálsson G and Prainsack B (2011) Genomic stuff: Governing the (im)matter of life. *International Journal of the Commons* 5(2): 259–283.

Pink S, Lanzeni D and Horst H (2018) Data anxieties: Finding trust in everyday digital mess. *Big Data & Society*. Epub ahead of print 31 January. DOI: 10.1177/2053951718756685.

Poovey M (1998) *A History of the Modern Fact: Problems of Knowledge in the Sciences of Wealth and Society*. Chicago, IL: University of Chicago Press.

Poulain M and Herm A (2013) Central population registers as a source of demographic statistics in Europe. *Population* 68(2): 183–212.

Raab CD (2016) Information privacy: Ethics and accountability. Available at: https://papers.ssrn.com/abstract=3057469 (accessed 20 February 2018).

Rapp R (2015) Big data, small kids: Medico-scientific, familial and advocacy visions of human brains. *BioSocieties* 11(3): 296–316.

Reiter JP (2002) Satisfying disclosure restrictions with synthetic data sets. *Journal of Official Statistics* 18(4): 531.

Rothstein B (2006) *Vad bör Staten Göra? Om Välfärdsstatens Moraliska och Politiska Logik*. Stockholm: SNS Förlag.

Ruckenstein M and Schüll ND (2017) The datafication of health. *Annual Review of Anthropology* 46(1): 261–278.

Sandell K, Berge E and Carlsson L (2003) The right of public access in Sweden: A history of modernization and a landscape perspective. In: Berge E and Carlsson L (eds) *Commons: Old and New* (NTNU ISS Rapport 70). Trondheim: Norwegian University of Science and Technology (NTNU), 49–58.

Seaver N (2017) Algorithms as culture: Some tactics for the ethnography of algorithmic systems. *Big Data & Society* 4(2). DOI: 10.1177/2053951717738104.

Seltzer W and Anderson M (2001) The dark side of numbers: The role of population data systems in human rights abuses. *Social Research* 68(2): 481–513.

Shaw SJ and Armin J (2011) The ethical self-fashioning of physicians and health care systems in culturally appropriate health care. *Culture, Medicine, and Psychiatry* 35(2): 236–261.

Shilton K (2013) Values levers: Building ethics into design. *Science, Technology, & Human Values* 38(3): 374–397.

Shore C (2008) Audit culture and illiberal governance: Universities and the politics of accountability. *Anthropological Theory* 8(3): 278–298.

Shore C, Wright S, Amit V, et al. (2015) Audit culture revisited: Rankings, ratings, and the reassembling of society. *Current Anthropology* 56(3): 421–444.

Sjöberg CM (2017) Scientific research and academic E-learning in light of the EU's legal framework for data protection. In: Corrales M, Fenwick M and Forgó N (eds) *New Technology, Big Data and the Law*. Singapore: Springer, 43–63.

Sköld P (2004) The birth of population statistics in Sweden. *The History of the Family* 9(1): 5–21.

Statens Offentliga Utredningar (SOU) (2017) Personuppgiftsbehandling för forskningsändamål. 9 June, Stockholm: Utbildningsdepartementet. SOU 2017: 50.

Steele JD (2002) Data protection: An opening door? The relationship between accessibility and privacy in Sweden in an EU perspective. *Liverpool Law Review* 24: 19–39.

Strathern M (2000a) Afterword: Accountability … and ethnography. In: Strathern M (ed.), *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. New York: Routledge, 279–304.

Strathern M (2000b) *Audit Cultures: Anthropological Studies in Accountability, Ethics, and the Academy*. New York: Routledge.

Sundin J, Hogstedt C, Lindberg J, et al. (eds.) (2005) *Svenska Folkets Hälsa i Historiskt Perspektiv* (Statens folkhälsoinstitut, 1651–8624 ; R 2005:8). Stockholm: Statens folkhälsoinstitut.

Tarkkala H and Tupasela A (2018) Shortcut to success? Negotiating genetic uniqueness in global biomedicine. *Social Studies of Science* 48(5): 740–761.

Tempini N and Leonelli S (2018) Concealment and discovery: The role of information security in biomedical data re-use. *Social Studies of Science* 48(4): 663–690.

Thomas R (2014) Accountability – A modern approach to regulating the 21st century data environment. In: Hijmans H and Kranenborg H (eds) *Data Protection Anno 2014: How to Restore Trust? Contributions in Honour of Peter Hustinx, European Data Protection Supervisor (2004-2014)*. Oxford: Oxford University Press, 135–147.

Tikkinen-Piri C, Rohunen A and Markkula J (2018) EU General Data Protection Regulation: Changes and implications for personal data collecting companies. *Computer Law & Security Review* 34(1): 134–153.

Tupasela A, Snell K and Cañada JA (2015) Constructing populations in biobanking. *Life Sciences, Society and Policy* 11(1): 5.

Varotto S (2015) The European General Data Protection Regulation and its potential impact on businesses: Some critical notes on the strengthened regime of accountability and the new sanctions. *Communications Law* 20(3): 78–85.

Venkat B (2017) Scenes of commitment. *Cultural Anthropology* 32(1): 93–116.

Versluis E (2007) Even rules, uneven practices: Opening the 'black box' of EU law in action. *West European Politics* 30(1): 50–67.

Veselý A (2013) Accountability in Central and Eastern Europe: Concept and reality. *International Review of Administrative Sciences* 79(2): 310–330.

Wagner J and Benecke A (2016) National legislation within the framework of the GDPR. *European Data Protection Law Review* 2: 353–361.

Wolters PTJ (2017) The security of personal data under the GDPR: A harmonized duty or a shared responsibility? *International Data Privacy Law* 7(3): 165–178.

Woolgar S and Neyland D (2013) *Mundane Governance: Ontology and Accountability*. Oxford: Oxford University Press.

Wuermeling U (1996) Harmonisation of European Union Privacy Law. *The John Marshall Journal of Information Technology & Privacy Law* 14(3): 411–460.

Zook M, Barocas S, boyd d, et al. (2017) Ten simple rules for responsible big data research. *PLoS Computational Biology* 13(3): e1005399.

Author biography

Alison Cool is an assistant professor in the Department of Anthropology at the University of Colorado Boulder. Her research, based on ethnographic fieldwork in Sweden, focuses on how experts, professionals, and activists go about the ethical and pragmatic work of protecting and sharing personal data. She is working on a book about data, personhood, and privacy.