



This is a repository copy of *The politics of data friction*.

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/120075/>

Version: Accepted Version

Article:

Bates, J. orcid.org/0000-0001-7266-8470 (2017) The politics of data friction. *Journal of Documentation*. ISSN 0022-0418

<https://doi.org/10.1108/JD-05-2017-0080>

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

The politics of data friction

Structured Abstract

Purpose

The aim of the paper is to further develop Paul Edwards' concept of "data friction" by examining the socio-material forces that are shaping data movements in the cases of research data and online communications data. The paper aims to articulate a politics of data friction, identifying the inter-related infrastructural, socio-cultural and regulatory dynamics of data friction, and how these are contributing to the constitution of social relations.

Design/methodology/approach

The article develops a hermeneutic review of the literature on socio-material factors influencing the movement of digital data between social actors in the cases of research data sharing and online communications data. Parallels between the two cases are identified and used to further develop understanding of the politics of "data friction" beyond the concept's current usage within the Science Studies literature.

Findings

A number of overarching parallels are identified relating to (1) the ways in which new data flows and the frictions that shape them bring social actors into new forms of relation with one another, (2) the platformisation of infrastructures for data circulation, and (3) state action to influence the dynamics of data movement. Moments and sites of "data friction" are identified as deeply political – resulting from the collective decisions of human actors who experience significantly different levels of empowerment with regard to shaping the overall outcome.

Research implications

The paper further develops Paul Edwards' concept of "data friction" beyond its current application in Science Studies. Analysis of the broader dynamics of data friction across different cases identifies a number of parallels that require further empirical examination and theorisation.

Practical implications

The observation that sites of data friction are deeply political has significant implications for all engaged in the practice and management of digital data production, circulation and use.

Social implications

It is argued that the concept of "data friction" can help social actors identify, examine and act upon some of the complex socio-material dynamics shaping emergent data movements across a variety of domains, and inform deliberation at all levels - from everyday practice to international regulation - about how such frictions can be collectively shaped towards the creation of more equitable and just societies.

Originality/value

The paper makes an original contribution to the literature on friction in the dynamics of digital data movement, arguing that in many cases data friction may be something to enable and foster, rather than overcome. It also brings together literature from diverse disciplinary fields to examine these frictional dynamics within two cases that have not previously been examined in relation to one another.

Introduction

The factors influencing the movement of digital data between social actors are increasingly being examined by research in Information Studies and related disciplines. The concept of “data friction” forwarded by Paul Edwards (2010) has been adopted by a number of scholars, primarily in the field of Science and Technology Studies (STS), to conceptualise some of the complex socio-material factors that coalesce to slow down and restrict data generation, movement and use. These studies have identified various sites of data friction that occur across a number of primarily scientific data infrastructures. However, they have not examined the wider politics of “data friction” as a social phenomenon. Often such studies are highly focused upon a particular disciplinary or interdisciplinary context, identifying and theorising the nature of specific forms of data friction and addressing how they might be overcome.

In this article, I shift focus in order to position “data friction” as being shaped by an emergent and complex politics of digital data movement. Rather than positioning “data friction” as necessarily problematic, I instead consider such frictions as something that are constituted within complex and contested socio-material spaces in which various forces struggle to shape how data do and do not move between different actors. In an era of “datafication” (Mayer-Schönberger & Cukier, 2013), it is important to understand how we might best advance knowledge using new forms of data, but also it is crucial to investigate the socio-material dynamics of how and why digital data do and do not move between actors with different, and at times conflicting, interests. Definitions of data are complex. Here, I place emphasis on the “computational” definition (Floridi, 2008), focusing on the socio-material dynamics influencing the electronic transmission of binary data. However, the underlying motivation is based upon an informational understanding of data as “alleged evidence” (Buckland, 1991), and questions and concerns around what these emergent digital data flows make visible and ‘knowable’ to who. Such insights will allow us to understand better how the circulation of different types of data contribute to the constitution of unfolding social relations. Taking such a perspective, it becomes clear that in many cases ‘data friction’ may be something to enable and foster, rather than something to overcome.

In this article, I synthesise insights from across the cross-disciplinary literatures that examine the socio-material factors that enable and restrict the movement of different types of data between social actors. I focus specifically on two sources of data around which significant efforts are underway in a number of countries to influence how data move between people and organisations: publicly funded research data and data generated from people's interactions with online communications platforms. These particular data sources were selected because 1) they involve complex data movements within and between key groups of social actors - citizens, science, state, market, 2) have been approached from different disciplinary perspectives, and 3) in the case of research data sharing there is a strong connection to the already existing work on data friction in scientific data infrastructures that is discussed in the following section.

My intention is not to produce an exhaustive systematic review of all barriers to data movement observed in the literature. Rather, I aim to develop a hermeneutic analysis of relevant literature from across the disciplines in order to identify key qualities of "data friction" in two particular contexts. I draw upon Kitchin's (2014) concept of a "data assemblage" to develop a framework through which to categorise three overarching factors identified within the literature as influencing "friction" in the movement of data: data sharing infrastructure and management, socio-cultural factors and regulatory frameworks. These factors are understood to be interrelated and developing in relation to broader political economic dynamics. From here, I consider how the socio-material frictions that restrict the movement of data between social actors can be understood as an important constitutive force in the development of social relations.

While research data and online communications data may initially appear to be quite different cases, a number of overarching parallels are identified relating to (1) the ways in which new data flows and the frictions that shape them bring social actors into new forms of relation with one another, (2) the platformisation of infrastructures for data circulation, and (3) state action to influence the dynamics of data movement. The concept of "data friction", I argue, can help social actors identify, examine and act upon some of the complex socio-material dynamics shaping emergent data movements across a variety of domains, and inform deliberation at all levels - from everyday practice to international regulation -

about how such frictions can be collectively shaped towards the creation of more equitable and just societies.

The social dynamics of digital data movement and friction

Clearly data do not move of their own accord. In order for “data friction” to exist there must be some force attempting to move data in the first place. Across the disciplines there have been various attempts to theorise and conceptualise the nature of how digital data move between social actors. A frequent framing emphasises the free flow of digital data across global networked infrastructure e.g. Castells (1991). This conceptualisation of ‘flowing’ data has, however, been challenged from a variety of perspectives; as Borgman (2015) observes it is clear that data don’t flow like oil. In the field of cultural theory, Scott Lash (2006) has emphasised the importance of “flux” – the tensions, struggles, and power dynamics that shape global information flows. Digital Sociologists such as Deborah Lupton (2014) have supported this argument, recognising the importance of understanding the “difficulties and blockages” in digital data flows, and David Beer (2013, p. 2) examines the ways digital data flow through popular culture, aiming to “locate an underlying politics of circulation” through examination of the data “flows, blockages and manipulations” and the impact of these circulations on the shaping of popular culture. Similarly, in geography there is deepening interest in the power dynamics of the emergent “global assemblage of digital flow’, and what this may mean for future socio-spatial relations (Graham, 2014; Pickren, 2016). Efforts to illuminate and theorise the socio-material constitution of these data movements are also articulated in my own recent work on ‘Data Journeys’ (Bates et al, 2016) and White’s (2017) related work on ‘Data Threads’. As White (2017, p. 93) argues, the notion of a data journey - as articulated in Bates et al (2016) - is particularly well attuned to addressing the “breaks, stoppages and disjunctures” that Edward’s concept of data friction alludes to.

While the above discussions tend to be relatively abstracted from everyday practices, within the information studies and STS literature we can observe more detailed empirical observations examining the nature of barriers to the movement of digital data and information across infrastructures and within organisations. For example, McNally et al. (2012) examine how “people, infrastructures, practices, things, knowledge and institutions” work together to shape the flow of data within data intensive research contexts, and

Leonelli (2013b) examines complex data integration issues in plant science. While scholars across disciplines have observed that there is a politics to digital data movements, it was from within this latter body of work on scientific data infrastructures that Edwards (2010) coined the term “data friction” - a term that has since been adopted by a number of STS researchers.

Edwards first developed the concept of “data friction” in his work on meteorological and climate knowledge infrastructures. In this initial articulation of the concept, he observes:

“Whereas computational friction expresses the struggle involved in transforming data into information and knowledge...data friction expresses a more primitive form of resistance – the costs in time, energy, and attention required simply to collect, check, store, move, receive, and access data. Whenever data travel – whether from one place on Earth to another, from one machine (or computer) to another, or from one medium (e.g. punch cards) to another (e.g. magnetic tape) – data friction impedes their movement” (Edwards, 2010, p. 84).

These frictions, he goes on to argue, are both social and physical in nature, and in the case of social systems “friction means conflict or disagreement” (p. 85). In this sense, friction is similar to the “flux” observed by Lash, however the notion of “data friction” is grounded in the identification of the materiality of digital data objects (Edwards, 2010, p. 84; Bates, Lin, & Goodale, 2016), and what might be described as a more empirically grounded and critical realist ontology (Edwards, 2010, p. 436-438).

In Edwards’ development of the concept, “data friction” is framed as a barrier to scientific advancement, and while friction can occur at any stage of the data lifecycle, emphasis is placed on how different forms of friction impede the movement and sharing of data between places, machines and mediums. Dawn Nafus (2014) further identifies the specific qualities of “friction” that lead to data becoming “stuck”. Similar to Edwards, she observes that some of these qualities are the result of the “numbers themselves”, while other factors are institutional and political in nature. Further, these qualities become visible at particular moments when data begin to move, pause, retreat and stop.

In further work, Edwards et al. (2011) go on to examine “data friction” in practice with a focus on how non-standardised metadata generation in a scientific collaboration resulted in

“data frictions” that impeded data sharing between disciplines, ultimately resulting in “science friction” across distributed e-science projects. As various scholars have noted, science studies has tended to focus on a variety of data problems within particular scientific disciplines; researchers have rarely examined how data travel between diverse disciplines (Edwards et al., 2011; Taylor et al., 2014). Further, we can observe that very little work in this tradition has examined how data travel between different groups of social actors and sectors. Bates et al.’s (2016) examination of the journeys of meteorological data between science, citizens, public and third sector organisations, and commercial organisations is one exception.

The concept of “data friction” has been adopted by a number of scholars, primarily within the STS tradition. For example, McCray (2014) draws upon the term to describe the tensions and technical challenges faced by astronomers as the data they work with became increasingly digitised. While most studies have focused on the physical sciences, a recent special issue of *Revue d'anthropologie des connaissances* edited by Jaton & Vinck (2016) examines “frictional moments” (p. c) observed in database projects in the Humanities and Social Science (HSS). They point to key tensions emerging around the status of HSS in the academy, the adequacy of digital tools, and the perceived credibility and utility of HSS as sources of friction in such projects. Beyond frictions in the circulation of research data, in her study of the Dutch land registry, Pelizza (2016) frames frictions as controversies about the best configurations of actors, agencies and sources to produce the most reliable data. She further observes that efforts to overcome friction at one site may simply displace frictions to another site, extending the circulation of data rather than removing friction altogether.

The existing literature on “data friction” has been centred on studies of scientific and research data infrastructures, and tends to be directed at understanding frictions in order to overcome them. This emphasis reflects Edwards et al.’s (2011) framing of the concept and call to “investigate how data traverse personal, institutional, and disciplinary divides” (p. 669). However, data friction is clearly a concept that translates beyond study of scientific data infrastructures. As has been addressed by scholars in the fields of digital sociology, geography, cultural studies and law, the movement of data between social actors, countries, platforms and so on, is a deepening force in the constitution of social relations. Efforts to

overcome data friction are also observed across industry. For example, Facebook CEO Mark Zuckerberg has stressed his desire to enable “frictionless sharing” for users (Payne, 2014). These efforts to move data are observed to influence, among other things, the production of popular culture (Beer, 2013), information streams and financial decisions (Pasquale, 2015), consumer desires (Turow, 2012), and citizen-state relations (Lyon, 2015), as well as the scientific knowledge emphasised in the above studies. While data friction may be a frustration to overcome in some cases for some people, in other cases, including those related to online activity, people feel powerless to generate enough friction in relation to the movement of data. It is clear, therefore, that the politics of data friction is more complex than currently conceptualised, and that a politics of data friction needs to consider not only sites of excess friction, but equally sites where friction is notable by its absence or lack.

Understanding the constitutive forces shaping data frictions

In this section of the paper, I synthesise insights from across the literature about drivers and challenges to the circulation of publicly funded research data and data that are captured as people interact with online communications platforms. In recent years significant questions have been raised about practices of data capture, sharing and publication in relation to each source of data e.g. through Open Science initiatives, surveillance practices, personalised marketing, etc.

As Kitchin (2014, p. 25) has argued, the complex and dynamic “assemblages” that produce data are made up of various interrelated elements, which I here adapt and breakdown into three analytical categories, each developing in relation to one another and the wider political economic dynamics of the market and finance:

- 1) Data sharing infrastructures and management (e.g. technical infrastructures, data management practices, organisations, materialities)
- 2) Socio-cultural factors (e.g. systems of thought, forms of knowledge, subjectivities, communities, institutions)
- 3) Regulatory frameworks (e.g. legalities, policy, standards).

Kitchin (2014, p. 24) observes that these various factors come together to frame what is “possible, desirable and expected of data” in different contexts. As demonstrated below,

these three categories also provide a useful analytical framework for illuminating the assemblage of socio-material factors at other stages of the data lifecycle, including those identified in the literature about what influences the dynamics of digital data movement.

Data friction in the circulation of publicly funded research data

The insights emerging from the STS and infrastructural studies literature about friction in scientific data infrastructures can be further developed through the lens of work on research data sharing and management. While academics have always engaged in practices of research data sharing across research teams, historically this was accomplished through personal networks and fostered through collegiality and trust (Sayogo & Pardo, 2013). While these informal methods of data sharing still exist, in recent years we can observe increased efforts to institutionalise and “normalize” research data sharing and re-use practices (Tenopir et al., 2015). The drivers for enabling data to move between social actors are various. For research data producers, increasing the reproducibility of research findings, and advancing science are key drivers. For data re-users, asking new questions of data and serving the “public interest” are additional concerns (Borgman, 2012). In the case of data licenced for re-use in non-academic contexts there is also the move to enable the exploitation of valuable publicly funded research data by commercial organisations, whether as open data or licenced by publishers (Murray-Rust, 2008). In some countries, significant public investments have been made to develop research data sharing infrastructures and platforms with the intention of fostering data sharing within disciplines e.g. (Leonelli, 2013a; Leonelli et al., 2013), and more widely. However, as Borgman (2012) observes, while data sharing is the norm in some disciplines e.g. genomics and astronomy, this is far from the case in many disciplines. As examined below, various socio-material challenges continue to be a significant source of “data friction” that hinder efforts to foster the movement of research data between different social actors.

Data sharing infrastructure and management

Despite infrastructural investments in data repositories and platforms, an increasing proportion of researchers sharing their data, and increased training and funding to support researchers’ data management practices, researchers report declining satisfaction with long-term data storage processes and tools for preparing the metadata needed to make

shared data re-usable by others (Tenopir et al., 2015). Potentially this is the result of increased expectations of researchers, who over the same time period also report an increased willingness to engage in data sharing. Nonetheless, it suggests that despite investment, the current data sharing infrastructure and data management practices still generate significant “friction” in efforts to move publicly funded research data between different social actors (Tenopir et al., 2011; Tenopir et al., 2015). These findings suggest that for many researchers, the data sharing infrastructure is “visible” (Star, 1999) and underdeveloped, rather than functioning seamlessly behind the scenes, and thus generates “friction” in the dynamics of research data circulation.

Data infrastructure developers face a multitude of challenges in enabling data to move between data producers and re-users. These include the complexity of scientific data (Koslow, 2002; Sayogo & Pardo, 2013), the unpredictability and dynamism of technological change (Bietz et al., 2016), the lack of standardised methods, data management and data sharing practices across disciplines (Reichman et al., 2011; Sayogo & Pardo, 2013; Borgman, 2012), the challenges of appropriately anonymising and sharing human subject data (King, 2011), and the barriers faced in financing sustainable open research data repositories (Kitchin et al, 2015). Examining some of these challenges in more depth, Leonelli (2013a) observes the immense challenges faced by database developers aiming to create a data sharing infrastructure for plant scientists studying the model plant *Arabidopsis thaliana* while respecting the diverse epistemic cultures within the discipline. She observes that the challenges did not stop at getting researchers to upload data to the database, but barriers were also experienced in making data in the repository accessible to potential re-users. In an attempt to overcome friction related to the re-usability of data across the discipline, developers attempted to curate the data in different ways for different epistemic communities by developing a variety of search interfaces that interrogated the database in different ways.

As Edwards et al. (2011) have observed, metadata practices are also a significant source of friction that restrict the re-usability of data, and thus limit its movement between different social actors. Without appropriate metadata, questions around database structures, quality and provenance arise for re-users (Elwood, 2008; Sayogo & Pardo, 2013), and datasets are unusable by third parties without significant direct communication between research teams

about how the dataset was constructed (Volk et al., 2014). It has been widely observed that a lack of quality metadata is a significant barrier to data sharing and re-use in various disciplines, making effective data re-use impossible in most cases (Alter & Vardigan, 2015; Volk et al., 2014; Reichman et al., 2011). Further, many researchers report that they do not have sufficient support or tools for preparing metadata so it can be re-used by third parties (Tenopir et al., 2015). Researchers' lack of skills, tools and motivation to create quality metadata are thus a significant source of friction for efficient data re-use that in many cases will mean that even if data are deposited they are still unlikely to travel beyond the repository.

In Tenopir et al.'s (2011) survey, "insufficient time" was perceived by researchers as the top reason for not sharing their data, a finding echoed by others (Volk et al., 2014; Alter & Vardigan, 2015). Relatedly, costs are also seen as prohibitive (Sayogo & Pardo, 2013). The political economy of research funding and the financial management of research institutions is clearly a critical issue that impacts on researchers' ability to engage in activities to reduce data friction. As mentioned above, data sharing is more common in disciplines where funding for research comes from public, rather than industry, sources (Borgman, 2012). However, as Alter & Vardigan (2015) and Kitchin et al (2015) observe, research funders and governments do not always cover the costs of data management and preservation, and while some countries devote funding to develop and sustain data sharing infrastructures, in other countries repositories are dependent on the uncertainty of grant funding cycles and a variety of non-public funding streams.

As Penny Andrews (2017) observes, publishers have begun to enter into this space, providing alternative services for data sharing. Data journals enable the publication of datasets in a similar way to how articles are published. However, as Murray-Rust (2008) observes, similar to articles, many publishers claim copyright on datasets restricting their re-use. A more recent development has been Elsevier's new research data sharing platform Mendeley Data that allows researchers to upload their own data similar to other academic and non-academic web-based 'sharing platforms'. Under this platform model, datasets are given a doi and depositing researchers are asked to choose between a number of Creative Commons and open licences in order to foster re-use of their data. While use of the

platform is currently free, a freemium business model is proposed for the future (<https://data.mendeley.com/faq>).

Socio-cultural factors

It has been observed by a number of researchers that while there is increasing agreement within the academic community with the principle of data sharing, what is currently possible and done in practice often diverges from this principle (Tenopir et al., 2011; Tenopir et al., 2015; Volk et al., 2014). Borgman (2012) observes that where advancing scientific research is the key underlying rationale for sharing data within a field e.g. in astronomy, the culture is more likely to be supportive of, and engaged in, mutually beneficial data sharing practices. However, significant data frictions can arise in data-dependent disciplines such as chemistry where data have a high monetary value and commercial interests mean results are often proprietary (Borgman, 2012). While disciplinary and national academic cultures clearly diverge in various ways, some general socio-cultural sources of “data friction” can be observed in the literature.

First are those sources of friction that cultural norms tend to perceive as appropriate restrictions on data movement, such as the desire to protect the confidentiality of participants, and restrict the sharing of identifiable or sensitive data without participants’ consent. In general, it can be observed that human subject researchers are significantly less likely to share their data than those in the physical sciences (Tenopir et al., 2015), in part due to such ethical concerns. However, there are other sources of friction that are perceived by some as legitimate, e.g. protection of intellectual property and potential for misinterpretation (Tenopir et al., 2015), that are less likely to be interpreted favourably by some within the open data advocacy community e.g. (Murray-Rust, 2008).

Second are those socio-cultural sources of friction which relate to the highly competitive environment that most academics and research teams are working within. Academic culture rewards researchers for their publications, not for their datasets (Sayogo & Pardo, 2013), and tends to incentivise self-interested behaviour. While there have been calls to adapt the citations and rewards system (Tenopir et al., 2015), there continue to be few incentives for researchers to publish data (Kaslow, 2002; Tenopir et al., 2015). While some researchers don’t recognise the value of their data for others (Tenopir et al., 2015), many do see the value and are wary of being “scooped” if they share data too early (Alter & Vardigan, 2015).

Many perceive a need to extract their own value from research data (i.e. through publishing findings) prior to sharing with others, and at a minimum, researchers tend to demand formal acknowledgement as an essential requirement of sharing (Tenopir et al., 2015). In particular, the literature indicates that these are concerns for those researchers whose position in the academic community is less secure – younger researchers (Tenopir et al., 2015) and those from lower and middle income countries with less access to the resources needed to quickly analyse and publish results (Jao et al., 2015). While often keen on data sharing in principle, in the context of contemporary academic culture such researchers may therefore generate data friction in order to protect their careers, research teams and national research standing from competitors (Tenopir et al., 2015; Jao et al., 2015).

Beyond these widespread competition concerns, some researchers may generate friction because they have concerns about subjecting their data and findings to public scrutiny (Tenopir et al., 2015). This may be an issue of particular concern in domains that attract significant public controversy such as climate science e.g. see Bates et al. (2016) and Edwards et al. (2011). However, researchers and research participants more generally report concerns around how data might be misinterpreted and misused by others (Alter & Vardigan, 2015; Jao et al., 2015; Tenopir et al., 2015). Relatedly, some researchers are also observed to experience a sense of ownership and responsibility for data they produce which results from their investment in time and resources. This sense of ownership, Jao et al. (2015) perceive, can result in a desire to influence how data is used by others and prevent misuse.

Regulatory factors

A variety of regulatory efforts to reduce the friction observed in the sharing of research data have emerged in recent years. Some of these developments mandate researchers to share data if, for example, they are in receipt of public funding or publish in a particular journal, and can therefore be perceived as pushing researchers' to find ways to overcome other barriers to data sharing in exchange for resources and prestige. Other regulatory frameworks address directly some of the concerns that researchers have about data sharing, and thus aim to shift perceptions of risk to individual researchers and teams.

Regulatory developments can be observed at various levels. For example, national funding councils are increasingly developing policies and recommendations that mandate or

encourage data sharing practices (Borgman, 2012). Tenopir et al. (2015) note such policies have been instituted by the National Science Foundation (USA), U.S. Office of Science and Technology Policy, Research Councils UK, Australian Research Council, and the European Commission. A number of scientific journals have also adopted data sharing and publication policies. Such policies either require or encourage authors to publish their datasets so that others can access and re-use them (Borgman, 2012). However, Tenopir et al. (2011) observed that authors are not always compliant and policy does not necessarily lead to datasets being available. Further, data availability does not necessarily mean re-usable data, if data are not licenced for re-use (Murray-Rust, 2008).

Institutional policies, Tenopir et al. (2011) observes, have “great influence on encouraging or inhibiting data sharing” by researchers. For example, some ethics policies can prohibit data sharing without participants’ consent (King, 2011), thus generating friction in data movements. Data management policies do not always encourage or mandate data sharing, however when they do, Sayogo & Pardo (2013) find that research data sharing policies that provide a strong framework to prevent poaching of data, ensuring acknowledgement of data producers, and assurance regarding misuse of data, are perceived positively by researchers considering data sharing. The existence of such a policy framework, they found, was a significant predictor of likelihood of researchers publishing datasets.

Beyond data sharing policies, the absence of metadata standards is frequently perceived as generating friction in data movements and there are various efforts underway to institute standards aimed at easing such frictions. For example, a recent JISC project in the UK has identified the development of “standard metadata profiles” as a key goal for increasing interoperability between research datasets (Kaye et al, 2017), and Zilinski et al (2016) stress the importance of research data metadata specialists in the development of standards for interoperability at an institutional, national and international level. As White (2017) discusses, data standards that define what is measured and how across different contexts are also increasingly used to produce indicators that enable interoperability and ease of comparison. However, these regulatory means of overcoming friction also raise challenges, as particular geographies and actors are prioritised when compromises and decisions are internalised into the structure of a standard (White, 2017).

Data friction in the circulation of online communications data

While the literature addressing the frictions that restrict the movement of publicly funded research data between different social actors tends to be concerned with overcoming barriers, the apparent ease with which data about people moves between organisations and sectors is the subject of increasing concern and, in some cases, controversy. The seeming lack of friction in the circulation of data captured about people's everyday interactions with online communications platforms such as social media, dating apps and messaging services, is one such area of concern. These often opaque movements of (potentially) identifiable and aggregated data work to varying degrees to make visible individuals and social groups (e.g. classified by demographics, personality, political beliefs etc) to actors across a variety of institutional contexts including government agencies, security and police services, employers, political campaigns, academic researchers, financial institutions and marketers.

Data sharing infrastructure and management

Technologies to enable the collection and circulation of data about people's online activity are baked into the infrastructure of the internet and web. Academic interest in these developments began to emerge in the late 1990s as researchers and digital rights advocates began to investigate the privacy implications of cookies which capture web users' browsing habits e.g. (Mayer-Schönberger, 1998). More recently, researchers have tended to focus on the infrastructural developments in the domains of state surveillance, platforms, APIs and the data brokerage industry.

Researchers have mapped the various, often opaque, ways in which platform users' communications data (both content and metadata) are shared with a variety of third party social actors. For example, APIs that connect third parties' systems with platform providers' databases (Vis, 2013); technologies such as social plugins that have extended platforms' data collection capabilities out to other websites; and, options for users to authenticate their identity on different sites using their social media accounts, enabling platforms to collect data about users' online activity (Sar & Al-Saggaf, 2013; Helmond, 2015; Plantin et al., 2016).

Complementing this socio-technical infrastructure, a new multi-billion dollar data brokerage sector has developed. Firms such as Acxiom and Datalogix collect and combine data from

various sources, enabling the aggregation and movement of vast streams of data about people between different actors. Particularly in the USA where legislation restricting personal data flows is relatively weak in relation to the EU, concerns have been raised about the significant lack of friction in this emergent market for (potentially) identifiable and aggregated personal data (Roderick, 2014; Pasquale, 2015). While the specifics of these infrastructural developments and data practices are not fully transparent, and in many cases are 'black-boxed' (Pasquale, 2015), it is clear that the capture and circulation of data about users is in many ways inseparable from the business models of the global internet and mobile communications infrastructures (Plantin et al., 2016).

Further, these circulations of data through the internet infrastructure are intercepted by a variety of actors, from criminal hackers to national security agencies. As evidenced by the documentation leaked by NSA contractor Edward Snowden, governments, national security services and private security companies around the world have invested millions of dollars in the development of sophisticated surveillance infrastructure through which digital communications and online activity can be intercepted and analysed (Lyon, 2015; Fuchs, 2013; Brown, 2014).

Clearly, the economic power of the primarily US-based technology firms behind many of these data generating platforms is a significant enabler of infrastructural developments to reduce friction in the circulation of such data. In an era where internet users' personal data fuels multi-billion dollar advertising and data brokerage industries (Roderick, 2014; Pasquale, 2015), it is clear that infrastructural investments aimed at overcoming data frictions that restrict firms' profits may come easier than in the case of publicly funded research data sharing infrastructure aimed at scientific collaboration and discovery.

While platform providers, data brokers and data interceptors aim to develop infrastructures that reduce friction in the collection and circulation of digital data, a variety of infrastructural technologies have also been developed that aim to counter these tendencies, from mainstream forms of encryption to less commonly used technologies such as the Tor browser. Technologies such as Virtual Private Networks which assign alternative IP addresses to computers, the Tor browser which enables anonymous web browsing, and DuckDuckGo which offers private search can all be understood as data friction generating technologies aimed at protecting users' privacy online (Brunton & Nissenbaum, 2015;

Macrina, 2015). Researchers and activists continue to develop new tools and technologies to help web users detect “privacy leakage” and take more control of friction in relation to their data e.g. (Ma, Meng, & Wang, 2012; Tomy & Pardede, 2016; Phillips, 2002).

Socio-cultural factors

While there have been year on year increases in usage of some “data friction” generating technologies such as DuckDuckGo (<https://en.wikipedia.org/wiki/DuckDuckGo>), the vast majority of internet users only take minimal precautions in relation to the capture and movement of their personal data when they go online.

For services such as Tor, network data suggests that use of Tor hidden services is largely confined to political activists and whistle-blowers, subcultures of hackers and privacy aware internet users, and people engaged in illegal activity (e.g. drug and gun markets, child abuse, pornography and counterfeiting) (Owen & Savage, 2015). While boyd & Hargittai (2010) found evidence that young people’s privacy awareness was increasing, and they were increasingly engaging in modification of Facebook privacy settings in order to increase friction in the circulation of their personal information, this is just one aspect of the online privacy jigsaw. There is a large body of research examining the socio-cultural factors influencing users’ perception and practices in relation to privacy online. Here we will focus on some of the literature that examines the socio-cultural dynamics that impact the moment of data production.

While some studies have found that users’ privacy concerns do not impact on their behaviour (Hsu, 2006), others have observed a more complex situation. For example, Drennan et al’s (2006) findings suggest privacy awareness does lead to increased suspicion about privacy risks and active user behaviour to generate friction. In relation to social demographic influences on behaviour, Sheehan (2002) observes that those with higher educational levels tend to be more concerned about online privacy, and that while older users tend to be either very concerned or not at all concerned, younger users tended to be more “pragmatic” in their decision making. Meanwhile, Krasnova et al (2010) found that people were motivated to disclose information online due to convenience of maintaining and developing relationships and platform enjoyment, a finding which contradicts O’Neil’s

(2001) earlier study that people prefer privacy over convenience. Further, Furnell et al. (2012) observe that in many cases while users are often concerned about risks, they are not fully informed, and therefore rather than protecting themselves they may decide to engage in what they understand to be risky behaviour.

These findings suggest that in response to the capture and circulation of data, there are a variety of complex user cultures emerging, potentially demographically differentiated, but which appear to have a tendency towards valuing social engagement, convenience and risk taking, tempered to varying degrees by basic friction generating behaviour such as avoiding particular types of information disclosure and taking action to modify platform privacy settings. However, while users may engage in practices that generate significant amounts of data that are then processed by platforms and, in some cases, distributed to third parties, research findings from the UK suggest that the majority are not happy about and do not actively consent to the data generated from their online activity being shared with and re-used by researchers, marketers and government (Evans et al., 2015). This discordance between peoples' practices and how they feel about how their data are used suggests significant disenfranchisement of internet users with regard to their collective ability to influence – and generate friction in – these data movements.

At the other end of these data journeys, at sites of data re-use, we can observe that the socio-cultural norms of people capturing and analysing data from social media platforms often pay little heed to these concerns. While there are diverse cultures of re-use, the cultural norm across much of industry, government and academia is that if it is publicly or legally available then the data can be ingested into databases to be processed and used for a variety of ends and paying little consideration to the perspectives of those whose activity is reflected in the data. Discourses of “publicness” and “openness” of data e.g. (Hoffmann et al., 2016), and legalistic references to privacy policies and terms and conditions which users agree to - even if they have difficulty understanding (McRobb, 2006) - help to reproduce a cultural norm that is enabling of data movement between platforms and third parties, and resists efforts to generate friction via critical reflection of practitioners and robust ethical frameworks. While professional organisations such as the ACM's Committee on Professional Ethics are in a process of updating ethics codes in response to these and wider concerns (see <https://ethics.acm.org>), there is currently little practical ethical guidance for data re-

users and it remains to be seen what impact such revisions will have on cultures of practice across different sectors.

Regulatory factors

While the infrastructural and socio-cultural conditions are increasingly enabling the circulation of data captured online these data movements are, to varying degrees, restrained by organisations' efforts to comply with friction-generating regulations on personal data sharing. Across jurisdictions, different regulatory approaches have been developed in an effort to shape data friction; however, enforcing compliance with regulation is a challenge faced by every country and there are many cases of organisations flouting regulatory restrictions on data movements, whether through negligence or in a purposeful attempt to gain competitive advantage. The challenges faced by states aiming to regulate the circulation of personal data are brought to the fore in the "complex and contentious" debates around the recently adopted EU General Data Protection Regulation, that were framed by political economic forces materialising in the powerful lobbying initiatives of industry, civil society and states institutions (Burri and Schär, 2016).

One significant area of concern for law makers has been the generation of friction in the movement of personal data across borders into different jurisdictions. Risks relating to the ease with which emerging internet technologies enable the movement of data across national borders, and thus into different data privacy regimes, have been a concern of regulators and academics since the 1980s as it became increasingly apparent that national law was struggling to govern international data flows e.g. (Rotenberg, 1994; Endeshaw, 1998). In an effort to generate the necessary border frictions to protect citizens' privacy within this emerging global information network, jurisdictions with stronger privacy regimes such as the European Union modified regulations in order to ensure that if citizens' data was moved beyond the border it would be afforded the same protections as within. In the case of the EU, as the infrastructure and business models have evolved, so too have the regulatory frameworks aimed at generating what the EU perceives to be an appropriate amount of data friction at its borders. In response to concerns raised about the NSA's access to EU citizens' data revealed by Snowden, in 2015 the European Court of Justice declared the 'Safe Harbour' agreement invalid. This agreement had allowed personal data transfer from the EU and US on the basis of firms' self-certification that they were protecting data in

line with EU law, and was replaced by a new set of requirements – called Privacy Shield – that aimed to increase friction in the movement of EU citizens’ data to the USA. However, former Secretary General to the French Data Protection Authority, Yann Padova (2016), questions the extent of the restrictions. He points out that the new rules will not restrict NSA access to data, nor protect against US authorities’ access to data that has been legally transferred to the US office of a multinational group of companies. Thus, the extent of the impact of the new rules on friction in the circulation of online communications data remain questionable.

Padova’s concerns regarding the NSA’s continued access to EU citizens’ personal data points to another development in the regulation of personal data flows post-Snowden: the development of laws on the interception and processing of citizens’ online communications data. While in the USA, the 2015 Freedom Act introduced new, friction generating, restrictions on US intelligence agencies collection of US citizens’ data, in the UK the Investigatory Powers Act of 2016 expanded the powers of UK intelligence and law enforcement agencies to carry out targeted and bulk interception of online communications data, thus reducing data frictions in the interception of UK citizens online communications and activity data.

Discussion: the politics of data friction

The above examination of some of the key factors shaping the dynamics of “data friction” in these two cases demonstrates that the socio-material dynamics constituting data friction are complex, influenced by a variety of infrastructural, socio-cultural and regulatory factors interrelated with the broader political economic context. The above also indicates that in order to understand what is happening at sites of data friction, it is important to observe not only the friction generating forces that are acting to restrict the movement of data, but also – and critically – the forces that are acting to move data between social actors. It is the relations between these differently directed forces that constitute the extent and nature of “data friction”. These dynamics of data friction are important because they influence what is made visible to and knowable by who, and therefore impact profoundly on the development of future knowledge and social relations.

The two sources of data considered in this paper are upon initial inspection quite different, and the literature identifies some seemingly conflicting socio-material forces influencing the dynamics of data friction in each case. For example, the desire for reproducibility and advancing research are key drivers for data sharing in the research context that appear to be substantially different from the clear profit drive observed in the case of online communications data. Poor quality metadata is a significant cause of friction in the circulation of research data, while metadata is lubricating the circulation of online communications data. Competition is identified as a source of friction in research data sharing, but a key driver for commercial actors to overcome some forms of friction in the circulation of online communications data. However, when we step back and consider the bigger picture there are a number of parallels that can help inform a deeper understanding of the politics of data friction across these different types of data.

The first parallel relates to how the movement of data in both cases brings together different, and often distant, social actors into new types of relations with one another. For example, the circulation of online communications data brings ordinary citizens into new forms of relationships with state agencies and commercial organisations that mine such data. Similarly, the circulation of research data brings scientific researchers into qualitatively new relations with one another and re-users of research data in different sectors. As Kitchin & Dodge (2011) theorised in relation to software code, we can also observe that the systems that enable and foster these data circulations also transform relations between people, objects, and so on. These data sharing infrastructures and platforms bring social actors into new forms of relation with one another, developers and owners. Through recognising these emergent forms of “data relations” (Kennedy, 2016), it becomes evident that the factors shaping data frictions are deeply political, and key to the struggle over how data mediates the relations between different groups of social actors.

As an example, we can observe that in both cases social and collaborative practices such as social interaction, resource sharing, enhancing collective understanding, and trust building are important drivers of data circulation. Yet, also we can observe a darker side to what this might mean in practice for differently situated groups of social actors. For example, the research literature suggests that those social actors with lower levels of security and/or power relative to others, for example younger researchers, researchers from low and

middle income countries, and ordinary internet users, appear to experience a tension between their desire to act sociably and the types of social relations they are resultantly drawn into. In the case of younger researchers embedded in increasingly competitive and time constrained academic culture, many respond to this tension by not sharing or delaying sharing data so they can avoid being “scooped” by competitors, a behaviour that goes against their somewhat more sociable and collaborative principles. A similar dynamic is observed in relation to online communications data. As it becomes increasingly difficult for people to disengage from online communications, young people’s efforts to manage online privacy concerns can be understood as an attempt to balance the tension between privacy enhancing behaviour and their desire to be engaged in social and collective life (Krasnova et al., 2010). This observation echoes findings about some publicly funded organisations’ concerns about opening data in the context of cuts to public funding and threatened services; the principle is supported, but the practice risks the potential for organisational harm due to the political economic context for data sharing (Bates et al., 2016). In all cases, people seemingly desire to engage in the forms of social and collective behaviour that are driving data circulation, and in many cases do despite risks to self. The friction arises largely from the increasingly competitive, market-driven social context in which this activity is compelled to take place - a context in which, as others have observed, powerful economic actors are highly dependent upon the exploitation of data generated from particular forms of collective life (Terranova, 2000; Scholz, 2012).

This observation leads to the second parallel between the two cases: the drive towards platformisation as a means to reduce friction and increase the circulation and exploitation of data. While over the last decade platform infrastructures and APIs have become a core strategy to enable data capture and reduce friction in the circulation of online communications data flows (Plantin et al., 2016), the research indicates that publicly funded research data infrastructures have struggled to become embedded into researchers’ practices. However, as Penny Andrews (2017) has observed we are beginning to see signs of a commercial take over and platformisation of research data sharing infrastructure with services such as Elsevier’s Mendeley Data launched in 2016. At the same time as these developments aim to foster data sharing and reduce data friction, Andrews (2017) argues they simultaneously deepen the trend towards sharing behaviour being absorbed into

commercial space aimed at generating profit from such practices. In an era when data is being lauded as the 'new oil', it is clear that firms such as Elsevier have a vested economic interest in accumulating high quality research data, and developing services around making that data available to and exploitable by others – whether re-users be academics or commercial actors.

The final parallel relates to the role of the state in the management of data friction. In both cases, we can observe regulatory action by states and commercial organisations to exert coercive force aimed at shaping how data move between social actors. This echoes wider regulatory developments in other areas such as open government data (Bates, 2014). While intellectual property restrictions clearly generate significant frictions in the circulation of some types of research data e.g. chemical data (frictions which benefit economically powerful industry actors), in many cases regulatory frameworks are increasingly acting as a coercive force to reduce friction in research data sharing. However, in countries such as the UK, mandates to share publicly funded research data emerge alongside other higher education policies which deepen the hyper-competitive, time pressured, and for many, insecure environment that research takes place within. Arguably, without addressing these wider cultural and economic issues, government mandates to share data add weight to the message that the neoliberal university should be ruled by a market logic, a logic that sees publicly funded data as a public good to be exploited by competing actors and therefore aims to restrict anti-competitive friction-generating practices by those with less security in that system. In this sense we can observe the promotion of cooperative sharing behaviour, without acknowledgement that such social practices are increasingly co-opted into the drive to deepen competition within, and marketization of, academic practice.

Meanwhile, in the case of online communications data we can observe neoliberal states and companies grappling with the complexities of maintaining a level of data friction that promises enough privacy and data protection to ensure the continued engagement of internet users, while simultaneously allowing enough circulation of data to allow it to be exploited by a variety of commercial, state and academic actors. The explicit consent clauses of the EU's new data protection regulations (GDPR) are a new development in relation to this dynamic, and it will be interesting to observe how the implementation of this regulation unfolds in the coming years. Relatedly, we can observe some states actively engaged in the

interception of online communications data are creating regulatory structures such as the UK's IPA that permit their activity in this area. While regulation currently does generate some friction in the circulation of potentially identifiable data, neoliberal state actors who advocate privacy experience a tension between protecting the rights of internet users and enabling the exploitation of their data by commercial actors and state agencies whose compliance with regulation is in most cases difficult to monitor and enforce. Thus while regulation is undoubtedly necessary in the generation of data friction, it will never be sufficient, and ultimately the responsibility for negotiating the tension between online activity and self-protective privacy is, in many cases, left to ordinary people engaging in everyday activities. As earlier interventions into the debate about technical versus regulatory solutions to online privacy concerns have addressed, it is therefore important to adopt a holistic approach when aiming to understand and address such challenges (Dourish & Anderson, 2006) (L. Edwards, 2003).

Conclusion

As the debates and struggles related to the shaping of digital societies unfold, it is becoming clear that "data friction" is central to many critical informational issues (e.g. open data, privacy, surveillance, data trading etc). As various actors work to make data move, a variety of socio-material counter forces potentially respond, generating "data friction" that slows and restricts data movements. Some of these forces have already been clearly observed in the research literature, while others identified in this article emerge from analysis of the broader dynamics of data friction across different cases. These broader dynamics of 'data friction' across different types of data circulation require further empirical examination and theorisation, as they point to the complex ways in which the constitution and implications of emergent data movements relate to the development of social relations, and how different social actors are differently positioned within this process.

Data friction influences what data are captured and how they are, or are not, made accessible and re-usable by different social actors, and ultimately how data movements are bringing social actors into new and complex forms of relation with one another. While originally developed in the context of understanding scientific data infrastructures, it is clear that the concept resonates with many contemporary debates and concerns about the production, distribution, processing and use of data across a variety of contexts, and has the

potential to inform a wider politics of data circulation. Through overcoming “data frictions” new informational practices are made possible and once hidden social and physical phenomena are made visible. Yet, in a world of deep social and economic inequalities it would be too simple to suggest that such developments impact all equally. As Fuchs (2011) observes, there are significant inequalities in who and what is being made transparent, and many of these new forms of data movement are enabling of forms of social management and surveillance that limit positive and political freedoms, and reproduce social inequalities. The above discussion demonstrates that the relationship between power and data friction is not simple, and that there are various examples of social actors with less power within contemporary structural conditions fostering “data friction” in an effort to retain or enhance their agency or position. Similarly, where data friction is lacking, we can observe that it is the result of similar human factors. This means that moments and sites of “data friction” are deeply political – they are the result of the collective decisions of human actors who experience significantly different levels of empowerment with regard to shaping the overall outcome. Such an observation has significant implications for all engaged in the practice and management of digital data production, circulation and use.

Alter, G. C., & Vardigan, M. (2015). Addressing Global Data Sharing Challenges. *Journal of Empirical Research on Human Research Ethics*, 10(3), 317-323.

doi:10.1177/1556264615591561

Andrews, P. (2017). *Foaming data: end to end and where will it end?* Paper presented at the International Labour Process Conference, Sheffield.

Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of Open Government Data in the United Kingdom.

Government Information Quarterly, 31(3), 388-395.

Bates, J., Lin, Y.-W., & Goodale, P. (2016). Data journeys: Capturing the socio-material constitution of data objects and flows. *Big Data & Society*, 3(2), 2053951716654502.

- Beer, D. (2013). *Popular culture and new media: The politics of circulation*. Basingstoke: Palgrave Macmillan.
- Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., . . . Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. *Journal of the American Medical Informatics Association*, 23(E1), E42-E48. doi:10.1093/jamia/ocv118
- Borgman, C. L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science and Technology*, 63(6), 1059-1078.
- Borgman, C. L. (2015). *Big data, little data, no data: scholarship in the networked world*. Cambridge, Mass.: MIT press.
- boyd, D., & Hargittai, E. (2010). Facebook privacy settings: Who cares? *First Monday*, 15(8).
- Brown, I. (2014). Social Media Surveillance. In *The International Encyclopedia of Digital Communication and Society* (pp. 1-7). Oxford: John Wiley & Sons, Inc.
- Brunton, F., & Nissenbaum, H. (2015). *Obfuscation: A user's guide for privacy and protest*. Cambridge, Mass.: MIT Press.
- Buckland, M. (1991). Information as Thing. *Journal of the American Society For Information Science*, 42(5), 351-360.
- Burri, M. & 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.
- Castells, M. (1991). *The informational city : information technology, economic restructuring, and the urban-regional process*. Oxford: Basil Blackwell.
- Dourish, P., & Anderson, K. (2006). Collective information practice: exploring privacy and security as social and cultural phenomena. *Human-computer interaction*, 21(3), 319-342.
- Drennan, J., Sullivan, G., & Previte, J. (2006). Privacy, risk perception, and expert online behavior: An exploratory study of household end users. *Journal of Organizational and End User Computing*, 18(1), 1-22.
- Edwards, L. (2003). Consumer privacy, on-line business and the internet: Looking for privacy in all the wrong places. *International Journal of Law and Information Technology*, 11(3), 226-250.

- Edwards, P. (2010). *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, Mass.: MIT Press.
- Edwards, P., Mayernik, M., Batcheller, A., Bowker, G., & Borgman, C. (2011). Science friction: Data, metadata, and collaboration. *Social Studies of Science*, 41(5), 667-690.
- Elwood, S. (2008). Grassroots groups as stakeholders in spatial data infrastructures: challenges and opportunities for local data development and sharing. *International Journal of Geographical Information Science*, 22(1), 71-90.
- Endeshaw, A. (1998). Regulating the Internet: clutching at a straw? *Computer Communications*, 20(16), 1519-1526.
- Evans, H., Ginnis, S., & Bartlett, J. (2015). *#SocialEthics: A guide to embedding ethics in social media research*. London: IPSOS-MORI. Retrieved 5th June 2016 from <https://www.ipsos-mori.com/Assets/Docs/Publications/im-demos-social-ethics-in-social-media-research-summary.pdf>
- Floridi, L. (2008). Data. In W. A. Darity (Ed.), *International Encyclopedia of the Social Sciences* (2nd ed.). Detroit: Macmillan.
- Fuchs, C. (2011). Towards an alternative concept of privacy. *Journal of Information, Communication and Ethics in Society*, 9(4), 220-237.
- Fuchs, C. (2013). Societal and ideological impacts of deep packet inspection Internet surveillance. *Information, Communication & Society*, 16(8), 1328-1359.
- Furnell, S., von Solms, R., & Phippen, A. (2012). Preventative actions for enhancing online protection and privacy. In Stowell, F (Ed.) *Systems Approach Applications for Developments in Information Technology*, Hershey, PA: IGI Global (pp. 226-236).
- Graham, S. (2014). Automated repair and backup systems. In N. Thrift, A. Tickell, S. Woolgar, & W. Rupp (Eds.), *Globalization in Practice*. Oxford: Oxford University Press.
- Helmond, A. (2015). The platformization of the web: Making web data platform ready. *Social Media+ Society*, 1(2), 1-11
- Hoffmann, A. L., Proferes, N., & Zimmer, M. (2016). "Making the world more open and connected": Mark Zuckerberg and the discursive construction of Facebook and its users. *New Media & Society*, 1-20.
- Hsu, C.-w. (2006). Privacy concerns, privacy practices and web site categories: Toward a situational paradigm. *Online Information Review*, 30(5), 569-586.

- Jao, I., Kombe, F., Mwalukore, S., Bull, S., Parker, M., Kamuya, D., . . . Marsh, V. (2015). Research Stakeholders' Views on Benefits and Challenges for Public Health Research Data Sharing in Kenya: The Importance of Trust and Social Relations. *Plos One*, *10*(9). doi:10.1371/journal.pone.0135545
- Jaton, F., & Vinck, D. (2016). Unfolding frictions in database projects. *Revue d'anthropologie des connaissances*, *11*(4), a-m.
- Kaye, J., Bruce, R., & Fripp, D. (2017). Establishing a shared research data service for UK universities. *Insights*, *30*(1), 59–70.
- Kennedy, H. (2016). *Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary*. Basingstoke: Palgrave Macmillan UK.
- King, G. (2011). Ensuring the Data-Rich Future of the Social Sciences. *Science*, *331*(6018), 719-721.
- Kitchin, R., Collins, S., & Frost, D. (2015). Funding models of Open Access digital data repositories. *Online Information Review*, *39*(5), 664-681.
- Kitchin, R. (2014). *The Data Revolution: big data, open data, data infrastructures and their consequences*. London: Sage.
- Kitchin, R., & Dodge, M. (2011). *Code/space: Software and Everyday Life*. Cambridge, Mass.: MIT Press.
- Koslow, S. H. (2002). Sharing primary data: a threat or asset to discovery? *Nature Reviews Neuroscience*, *3*(4), 311-313.
- Krasnova, H., Spiekermann, S., Koroleva, K., & Hildebrand, T. (2010). Online social networks: why we disclose. *Journal of Information Technology*, *25*(2), 109-125.
- Lash, S. (2006). Life (Vitalism). *Theory, Culture & Society*, *23*(2-3), 323-329.
- Leonelli, S. (2013a). Global data for local science: Assessing the scale of data infrastructures in biological and biomedical research. *BioSocieties*, *8*(4), 449-465.
- Leonelli, S. (2013b). Integrating data to acquire new knowledge: Three modes of integration in plant science. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, *44*(4), 503-514.
- Leonelli, S., Smirnoff, N., Moore, J., Cook, C., & Bastow, R. (2013). Making open data work for plant scientists. *Journal of Experimental Botany*, *64*(14), 4109-4117. doi:10.1093/jxb/ert273
- Lupton, D. (2014). *Digital sociology*. London: Routledge.

- Lyon, D. (2015). The Snowden stakes: challenges for understanding surveillance today. *Surveillance & Society, 13*(2), 139-152.
- Ma, R., Meng, X., & Wang, Z. (2012). Preserving privacy on the searchable internet. *International Journal of Web Information Systems, 8*(3), 322-344.
- Macrina, A. (2015). Accidental Technologist: The Tor Browser and Intellectual Freedom in the Digital Age. *Reference & User Services Quarterly, 54*(4), 17-20.
- Mayer-Schönberger, V. (1998). The Internet and Privacy Legislation: Cookies for a Treat? *Computer Law & Security Review, 14*(3), 166-174.
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: A revolution that will transform how we live, work, and think*. London: Houghton Mifflin Harcourt.
- McCray, W. P. (2014). How Astronomers Digitized the Sky. *Technology and Culture, 55*(4), 908-944.
- McNally, R., Mackenzie, A., Hui, A., & Tomomitsu, J. (2012). Understanding the 'intensive' in 'data intensive research': Data flows in next generation sequencing and environmental networked sensors. *International Journal of Digital Curation, 7*(1), 81-94.
- McRobb, S. (2006). Let's agree to differ: varying interpretations of online privacy policies. *Journal of Information, Communication and Ethics in Society, 4*(4), 215-228.
- Murray-Rust, P. (2008). Open data in science. *Serials Review, 34*(1), 52-64.
- Nafus, D. (2014). Stuck data, dead data, and disloyal data: the stops and starts in making numbers into social practices. *Distinktion: Journal of Social Theory, 15*(2), 208-222.
- Owen, G., & Savage, N. (2015). *The Tor Dark Net*. Global Commission on Internet Governance paper series no. 20. Retrieved 1st May 2017 from https://www.ourinternet.org/sites/default/files/publications/no20_0.pdf
- O'Neil, D. (2001). Analysis of Internet users' level of online privacy concerns. *Social Science Computer Review, 19*(1), 17-31.
- Padova, Y. (2016). The Safe Harbour is invalid: what tools remain for data transfers and what comes next? *International Data Privacy Law, 6*(2), 139-161.
- Pasquale, F. (2015). *The black box society: The secret algorithms that control money and information*. Cambridge, Mass.: Harvard University Press.
- Payne, R. (2014). Frictionless Sharing and Digital Promiscuity. *Communication and Critical/Cultural Studies, 11*(2), 85-102.

- Pelizza, A. (2016). Disciplining Change, Displacing Frictions. Two Structural Dimensions of Digital Circulation Across Land Registry Database Integration. *Tecnoscienza: Italian Journal of Science and Technology Studies*, 7(2), 35-60.
- Phillips, D. J. (2002). Negotiating the digital closet: Online pseudonymity and the politics of sexual identity. *Information, Communication & Society*, 5(3), 406-424.
- Pickren, G. (2016). 'The global assemblage of digital flow' Critical data studies and the infrastructures of computing. *Progress in Human Geography*, OnlineFirst Retrieved from: <http://journals.sagepub.com/doi/abs/10.1177/0309132516673241>.
- Plantin, J.-C., Lagoze, C., Edwards, P. N., & Sandvig, C. (2016). Infrastructure studies meet platform studies in the age of Google and Facebook. *New Media & Society*, OnlineFirst, 1-18
- Reichman, O. J., Jones, M. B., & Schildhauer, M. P. (2011). Challenges and Opportunities of Open Data in Ecology. *Science*, 331(6018), 703-705.
- Roderick, L. (2014). Discipline and Power in the Digital Age: The Case of the US Consumer Data Broker Industry. *Critical Sociology*, 40(5), 729-746.
- Rotenberg, M. (1994). Electronic privacy legislation in the United States. *The Journal of Academic Librarianship*, 20(4), 227-230.
- Sar, R. K., & Al-Saggaf, Y. (2013). Propagation of unintentionally shared information and online tracking. *First Monday*, 18(6). Retrieved from: <http://firstmonday.org/article/view/4349/3681>
- Sayogo, D. S., & Pardo, T. A. (2013). Exploring the determinants of scientific data sharing: Understanding the motivation to publish research data. *Government Information Quarterly*, 30, S19-S31.
- Scholz, T. (2012). *Digital labor: The Internet as playground and factory*. New York: Routledge.
- Sheehan, K. B. (2002). Toward a typology of Internet users and online privacy concerns. *The Information Society*, 18(1), 21-32.
- Star, S. (1999). The ethnography of infrastructure. *American Behavioral Scientist*, 43(3), 377-391.
- Taylor, A., Fisher, J., Cook, B., Ishtiaq, S., & Piterman, N. (2014). Modelling biology – working through (in-)stabilities and frictions. *Computational Culture: a journal of software studies*. Retrieved from: <http://computationalculture.net/article/modelling-biology>

- Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., . . . Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *Plos One*, 6(6). doi:10.1371/journal.pone.0021101
- Tenopir, C., Dalton, E. D., Allard, S., Frame, M., Pjesivac, I., Birch, B., . . . Dorsett, K. (2015). Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. *Plos One*, 10(8). doi:10.1371/journal.pone.0134826
- Terranova, T. (2000). Free labor: Producing culture for the digital economy. *Social text*, 18(2), 33-58.
- Tomy, S., & Pardede, E. (2016). Controlling privacy disclosure of third party applications in online social networks. *International Journal of Web Information Systems*, 12(2), 215-241.
- Turow, J. (2012). *The daily you: How the new advertising industry is defining your identity and your worth*. New Haven, CT.: Yale University Press.
- Vis, F. (2013). A critical reflection on Big Data: Considering APIs, researchers and tools as data makers. *First Monday*, 18(10).
- Volk, C. J., Lucero, Y., & Barnas, K. (2014). Why is Data Sharing in Collaborative Natural Resource Efforts so Hard and What can We Do to Improve it? *Environmental Management*, 53(5), 883-893.
- White, J. M. (2017). Following data threads. In R. Kitchin, T.P. Lauriault, & G. McArdle (Eds.), *Data and the City*. [Abingdon](#): Routledge, pp. 85-97.
- Zilinski, L. D., Barton, A., Zhang, T., Pouchard, L., & Pascuzzi, P. (2016). Research data integration in the purdue libraries. *Bulletin of the American Society for Information Science and Technology*, 42(2), 33-37.