Commentary on Part I

'The Hairs of Hope': Legal and Material Infrastructures in a Global Context

MARIANA VALVERDE

I am not an expert on either digital flows or infrastructures: I am a social-political theorist, turned socio-legal theorist, who like many of us is attempting to deal, intellectually and politically, with the information practices of the present. In my recent work I have paid particular attention to information practices and representational strategies in the context of major infrastructure projects, but without specific attention to software or platform mediated practices.

Therefore in this comment I will address, somewhat tangentially, some of the points made in the two initial chapters of this book by Julie Cohen and Jen Raso; but I will mainly focus on questions that pertain to the larger enterprise, that is the research project that has given rise to this book.

Scanning much of the literature on 'big data', as well as the almost completely separate literature on infrastructures,[1] it has long struck me that many key terms are taken for granted even in critical analyses. Thus, while in my youth I studied what was called 'continental' philosophy,[2]

[1] The literature on 'smart cities' often combines attention to infrastructure governance with attention to 'big data'. However, most scholars in this subfield come either from 'big data' and surveillance studies or else from the separate realm of urban studies/infrastructure, and hence tend to adopt research questions already raised in their preferred fields. Stephen Graham, Simon Marvin, Andres Luque Ayala and Colin McFarlane are influential scholars located within urban studies and infrastructure studies, and from that perspective they have made excellent contributions to the study of what Julie Cohen calls 'the datafication of infrastructure'. Unfortunately their work has not usually circulated within legal circles, in which issues of 'datafication' have been mainly raised from the perspective of intellectual property law or that of privacy law.

[2] That 'the continent', in English usage, refers almost solely to France and Germany, is of course a great example of lingering Eurocentrism. But today's philosophy departments,

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

not analytic (English) philosophy, I believe that the kinds of questions most often asked by analytic philosophers (such as: 'what do you mean by 'data' anyway?') can be very useful to clear the ground to prepare it for discussions such as the ones represented in this volume. This comment is thus meant as a contribution to the collective discussion rather than as the kind of 'comment' on individual papers usually presented in academic conference panels.

In highlighting what is unique about our present, from a progressive perspective – which in my view means that we avoid presupposing that capitalist market mechanisms are usually better than alternatives – most people turn to words such as 'the information society' or 'big data-society'. These terms have given rise to derivative terms – perhaps most influentially, the term 'data-driven', a term used as it usually is in Julie Cohen's contribution. I find the very term 'data-driven' problematic, in that the recourse to data, big or otherwise, is mainly a means to an end – not the kind of Aristotelian end that drives processes in a teleological manner. Data infrastructures can be means to a public-purpose end, but the same tools can also serve the goal of increasing market share and/or profits for a capitalist corporation. Of course it is necessary for public institutions to pay attention to the often unintended effects of 'datafication'. This scrutiny is not easy since digital tools most often come as a blackboxed package (as when Canadian universities, in the early twenty-first century, chose either Microsoft or Google for software services that become compulsory for students and staff, with little scrutiny of the effects of these choices).

Perhaps because I am a newcomer to the field of 'big data' studies, I hold that people and institutions are driven by certain pre-digital, old-fashioned motivations: to increase profits and/or market share, in the case of for-profit corporations, to promote democracy and accessibility in the case of NGOs, or to increase or modify social control in the case of authoritarian states.

This raises a broader point about the way in which we study data flows. While as a Foucaultian of sorts I do pay particular attention to 'how' questions, which in scholarly practice means mainly analysing the effects of particular techniques of governance,[3] I have come to believe that it may be risky, both politically and intellectually, to do research as if the

even in North America, tend to take the English meaning of 'the continent' for granted. Russia, Spain, Greece and Italy are not really 'continental' in philosophy department speak.

[3] An example from the realm of universities is that the conditions contained in the contracts signed between 'the university' (often, the head of IT) and providers of software packages

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

forms of power and the character of the networks made possible by modern computing are wholly new. I hold that while we may live in data-saturated fields of governance, the moves made by authorities and would-be authorities that implement digital systems are not necessarily wholly new. There may be more continuity between the pre-digital past and the present than is often assumed by writers who focus on the innovative features of the 'information' or 'big data' society.

Jen Raso's brilliant chapter is a case in point, in my view. The dialectical dynamic she attributes to digital processes utilised in many state agencies (a dynamic that features simultaneous processes of centralisation and dispersal) is, I would argue, far older than the modern computer. My research on the history of special-purpose districts and of arms-length semi-public agencies governed not through ministerial accountability but by means of relatively autonomous corporate-like boards has found that the dispersal of state 'agency' through special-purpose bodies and what British administrative lawyers call 'the boardisation of the state' coexists with a kind of rebirth of sovereign power: the arms-length agencies, Quangos and so on, operate as if they were private for-profit corporations, subjecting people and entities to top-down undemocratic rule.[4] Under the populist banner of 'depolitisation', starting well over a hundred years ago, many public and quasi-public agencies have been turned into what one might call fake private corporations. These have minimal transparency requirements and even less by way of accountability mechanisms. This whole complex 'ecosystem' of governance tends to be neglected by political science, and its unique features are not necessarily captured by administrative law's category of 'agencies, boards and commissions'.

are not made public, even within the institution. Students and staff are obligated to use tools such as Microsoft Teams and Outlook, or Google Meet and Gmail, and are not asked to provide even the token 'I accept the terms and conditions' that one clicks on one's phone when downloading an app. The fact that in Canada virtually all universities are public does not in itself ensure that there's input, even from the internal publics, into software choices. In infrastructural terms, universities are nearly all publicly owned, but as we know from the realm of transportation, the formal ownership of an infrastructure is not necessarily an indicator of how key decisions are made.

[4] See M Valverde, 'Ad Hoc Governance: Public Authorities and North American Local Infrastructure in Historical Perspective' in M Brady and R Lippert (eds), *Governing Practices: Neoliberalism, Governmentality and the Ethnographic Imaginary* (University of Toronto Press 2018); A Flynn and M Valverde, 'Where the Sidewalk Ends' (2019) 36 *Windsor Yearbook of Access to Justice* 263; M Valverde and A Flynn (eds), *Smart Cities in Canada: Digital Dreams, Corporate Designs* (Lorimer 2021), especially the Introduction, and M Valverde, *Infrastructure: New Trajectories in Law* (Routledge 2022).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

A significant literature relied upon in this volume concerns the processes by which corporations (especially oligopolistic ones based in either Silicon Valley or, as Julie Cohen helpfully reminds us, China) appropriate the 'data exhaust' emitted by us and our 'devices' as we work, seek health care, shop for goods, engage in leisure activities, educate our children or travel. That the process by which private corporations (be they for-profit US corporations or corporations controlled by the Chinese party-state) appropriate the vast reams of what we have learned to call 'data' generated both by individuals and by infrastructures in use as well as entities such as Jen Raso's 'state agencies' has come to be known as 'data extractivism' – through a mostly implicit parallel with the better known older practice of the state licensing for-profit companies so they can legally appropriate arguably common natural resources such as timber, water, and minerals. Data extractivism is a useful term, in highlighting the monetary value of data that has no value for the individual (I cannot bring to any market I know the location data generated by my phone) but has much value for platform businesses such as Google, Airbnb, Amazon and so on, and also for commercial data brokers.

Data extractivism is a very useful term in discussions about the nature and the likely future of what Shoshana Zuboff has influentially called 'surveillance capitalism', in part because it underlines the commonalities between the management of 'physical' resources whose enclosure and appropriation has long been studied and today's digital or electronic resources. Fortunately, this book focuses not so much on the kind of data extractivism that privacy activists fight against, but rather on broader patterns of data extractivism through which both states and corporations build networks based on the extraction of intangible resources (including but not limited to datasets, or aggregate data not yet made into a dataset).

Personal privacy gets much public attention (e.g. the European Union's 'right to be forgotten'); but as many have pointed out, that battle was arguably lost when Google was founded, and in any case impersonal aggregate data has been found to be an excellent resource for many corporations. The ubiquitous use of assemblages of digital information, software and hardware such as the microchips embedded in 'ordinary' objects can be said to have infrastructural effects, as Julie Cohen points out in her chapter. This is very true, but what does that claim mean, exactly? It seems to me that we collectively need to have a little bit of clarity about what the highly mobile and currently ambiguous term 'infrastructure' does. There is much talk in the social sciences (not so

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

much in law) about the 'infrastructural turn', but little clarity about what that might mean beyond trying to redefine and resignify many entities and processes under the banner of 'infrastructure'.

The two key terms that need reflection and commentary, for this volume as a whole, are 'data' and 'infrastructure'. I will begin with the less familiar (in the law and technology field) term of 'infrastructure', since I have carried out empirical research on the governance of infrastructure projects.[5] I will then jump into the pool – or swamp? – of digital-world discussions by asking a few questions about that largely taken for granted term, 'data'.

What Do We Mean by 'Infrastructure' and 'Infrastructural Effects'?

Infrastructure is traditionally defined as networks that provide public services, not necessarily through state ownership and control but to meet public as distinct from individual needs: communication, transportation, the energy to power things and activities, the management of both waste water and drinking water and so on. The term has no fixed boundaries, since in the global South or majority world what might look like 'jerry-rigging' might work reasonably well as an infrastructure.[6]

But in this day and age, infrastructure is seen as more than physical. That is not a result, it needs to be highlighted, of the intangible character of digital data. Well before the internet, telephone communication was enabled by cables and phones but the importance of the content of the communications could not be reduced to gadgets. And conversely, despite the tendency in legal circles for digital data to be regarded as purely

[5] M Valverde and A Moore, 'The Performance of Transparency in Public-Private Infrastructure Project Governance: The Politics of Documentary Practices' (2019) 56(4) *Urban Studies* 689; M Valverde, F Johns and J Raso, 'Governing Infrastructure in the Age of "the Art of the Deal": Logics of Governance and Scales of Visibility' (2018) 41 *Political and Legal Anthropology Review* 118.

[6] Some have argued that there is a crucial qualitative difference between the 'modern infrastructure ideal' of the global North, featuring integrated systems that work with little internal friction, and the more inventive and not always legal mechanisms that fulfil, well or badly, similar needs in the global South. I prefer to avoid the North-South binary, in line with the idea of 'ordinary cities' first presented by Jennifer Robinson (cite) and developed by many other scholars. See, for example, S Graham (ed), *Disrupted Cities: When Infrastructure Fails* (Routledge 2009) and the highly influential work in S Graham and S Marvin, *Splintering Urbanisms: Networked Infrastructures, Technological Mobilities and the Urban Condition* (Routledge 2001). See also S Graham, 'Splintering Urbanism at 20' (2022) 1 *Journal of Urban Technology* 169.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

'intangible' (perhaps because of the prominence and dissemination of intellectual property (IP) terms of art), the datasets cannot exist in thin air: they too require cables (the submarine cables that make the internet global being often literally attached to existing phone cables) and of course physical computers, many of them quietly gathered by big corporations in anonymous-looking hangar-like server farms.[7] That is: what are often dismissively called 'legacy' infrastructures include both physical and intangible components, and so too does the world of digital data. Legal scholars tend to underestimate the materiality of law,[8] as many critical socio-legal scholars have noted, so it is not surprising that they have also tended to underestimate the materiality of digital data – especially since IP law insists on classifying property as either 'tangible' or 'intangible', an oversimplification of complex realities that is parallel to and perhaps based on the assumed binary distinction between 'real' and 'moveable' property.

Like Foucault pondering what came before 'the deployment of sexuality', if we want to understand the proliferation of the term 'infrastructure' we need to consider what it replaced, what came before. The answer is easy: in countries around the world ministries of infrastructure replaced what were until recently ministries of public works. Municipalities tend to still have a 'public works department'; but at higher levels of government what we have are infrastructure ministries and, even more important, infrastructure 'agencies' (to re-use Jen Raso's useful term, recovered from the boring public administration literature on bureaucracies). Why the change?

I argue that the varied phenomena that huddle under the ambiguous megaterm 'neoliberalism' have combined to create a devil-like negative halo around public expenditures – and hence public works. But of course roads need to be built and maintained today as in the past, as do water systems, electricity grids, railway networks and so on. So what was the solution to this problem of justifying expenditures on public works within neoliberal regimes? The solution that was palatable to neoliberal politicians was a semantic one: to rename the public expenditures

[7] A useful resource in this regard is the popular-science book by A Blum, *Tubes: Journey to the Center of the Internet* (Harper Collins 2012).

[8] Some legal scholars, mainly in the UK and Australia, have been highlighting the materiality of law for decades, but in recent years the most important reference point has been Bruno Latour's ethnography of the Conseil d'Etat, published in English as B Latour, *The Making of Law* (M Brilman and A Pottage trs; Polity 2009). However, the original French title *La Fabrique du Droit* (La Decouverte 2002) has rich associations lost in translation.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

previously listed under ministries of public works as 'investments'. The operative phrase became 'investing in infrastructure'. Try googling that phrase for the jurisdiction you live in and you'll be persuaded by my argument.

Just as public expenditures could be saved for neoliberal purposes by being renamed 'investments',[9] so too the term infrastructure, which as many scholars have pointed out does not have a single clear meaning, was highly useful. Even the most right-wing tax-cutting politicians can boast about 'investing in infrastructure' – and they do, regularly. Even Boris Johnson, in his day, attempted to show that he too was 'investing in infrastructure' by financing with huge amounts of public money the ill-fated British high-speed rail project known as HS2. In the meantime, social democrats, and even the Chinese communist party, are also incentivised to adopt the new language of investing in infrastructure to justify pursuing New Deal-style dreams of major public works. And, relevant to the expansiveness of the term 'infrastructure' in today's world, social democrats usually add that by infrastructure they don't just mean roads and bridges and pipelines; they are also in favour of investing in 'social infrastructure' (formerly known as the welfare state).

It seems that virtually everything can be refigured as infrastructure in the present day. If Raymond Williams were to rise from his grave to analyse some 'keywords' of our own present, he'd have to include 'infrastructure'. Re-naming welfare state projects such as homes for the aged or daycares for toddlers as 'social infrastructure' serves to justify spending public money on them at a time when 'public expenditure' is a no-no word.

The American pragmatists turn out to have been right: 'infrastructure' is whatever the term does in practice.[10] One can even redefine parts of the environment as infrastructure, if one stresses the public or social value of entities such as lakes, forests and beaches, whether or not those are in fact publicly owned. As has been pointed out by many infrastructure experts, the question of who owns a resource is not always determinative.

[9] Investments, in capitalist economics, are meant to generate future profits. In relation to many if not most infrastructure projects the expenditures do not result in future profits, and were never meant to and so the term 'investing' is inaccurately used; but that doesn't seem to bother any of the myriad politicians using it.

[10] The most famous statement of the pragmatists' theory of truth was that by William James to the effect that God exists if and insofar as humans' belief in God has real-world effects. That kind of argument sidelines metaphysical arguments about whether or not entities including God 'really' exist or not.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

A railway can serve the public at large, as a legally enforceable obligation, even if it is privately owned (as in the US legal doctrine of the responsibilities of 'common carriers').

Brett Frischmann is a noted US legal scholar who has argued that the term 'infrastructure' has an inherently public or at least shared logic. Like most American legal scholars he is keener on empowering civil society than in expanding the reach of the state; but in arguing that infrastructural goods are those that are not 'enclosed' but rather can be used by an indefinite number of people and groups for diverse purposes, he seems to line up, at least in some respects, with the social democratic notion first mooted (in American legal context) by noted property scholar Carol Rose. In 1986 Carol Rose published an influential article, in the *Chicago Law Review* of all places, on 'the comedy of the commons' – meaning the opposite of the better known 'tragedy of the commons'.[11] In that article she presented the idea that there are some entities that are 'inherently public' property. That claim has certainly been challenged, especially by those that would make a distinction between 'state owned and controlled' on the one hand and 'public' on the other. But for present purposes we need not take a stand on the difficult question of whether state property is or is not coterminous with public property – or whether the category of 'the commons' is a useful one ('the digital commons' being a term of highly uncertain content).

Here let me briefly try to address Jen Raso's and Julie Cohen's chapters together by focusing on Raso's claim that datafication (or maybe just some forms of software use?) is a process studied by critical scholars only as it negatively impacts individuals (through breaches of privacy) or on groups (through the discrimination embedded in what AI people call 'training data').[12] Raso shows, through secondary literatures and through the detailed study of the case of the hapless Algerian denied entry into Canada to visit his brother, that the systematic and use of networked data has a constitutive effect on 'the state'. Instead of adding to the by now stale discussions about the character and extent of 'the state', however, Raso usefully highlights the dual role of state agencies (known in other literatures as bureaucracies, a term that is also intrinsically ambiguous).

[11] C Rose, 'The Comedy of the Commons: Custom, Commerce, and Inherently Public Property' (1986) 53 *University of Chicago Law Review* 711.

[12] The very term 'training data' could and probably should be subject to critical analysis, since in many ways all datasets can and perhaps do become 'training' data, but here that would be a digression.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

On the one hand, sometimes collaborating and sometimes competing state agencies utilise digital datasets not only to do their own work but to communicate with other agencies, across ministries and even across national boundaries. Thus, on the other hand, the supposed unified agenda and 'agency' of the state becomes fragmented and dispersed. This dispersal means that legal mechanisms designed to address complaints and provide institutional accountability become moot or useless: it is very difficult, even impossible, for citizens to clearly identify which agency treated them inappropriately, or for journalists and scholars to find a single culprit to denounce.

Raso's focus on specific state agencies is a breath of fresh air, after decades of largely unproductive theoretical debates about 'the state'. (These debates, younger scholars may not know, petered out around the fall of the Berlin wall, as progressives rushed to embrace 'civil society' and sidelined the inconclusive intra-Marxist debates about 'the state'.)

However, one point I would make, not so much by way of critique as by way of suggestions for further research, is that the strange dialectic whereby state power seems to concentrate and disperse at the same time is hardly new. The use of interconnected datasets – adopting here Rob Kitchin's definition of 'big data', which stresses mobility and connectivity of datasets rather than simply size[13] – can certainly facilitate the dispersal of accountability across agencies, while simultaneously bringing back the oldest mechanisms of sovereign state power. But this dialectic of centralization and dispersion is visible in pre-internet, pre-digital state systems, if one sets aside the Marxist spectacles through which 'the state' was presumed to be unified by its being based on the interests of the ruling class.

Administrative law's imaginary features particular state agencies acting autonomously, which as Raso points out can be problematic for mistreated citizens who may be the victims of assemblages of data exchanges and data flows. But state agencies have for a long time now had both the ability and the incentive to hive off subsidiaries and otherwise foster wholly new entities. Agencies as well as legislatures and ministries have long engaged in the co-creation, either with other parts of government or with the private sector, of new agencies – especially 'arms-length' special-purpose agencies.

[13] R Kitchin, *The Data Revolution: Big Data, Open Data, Data Infrastructures and Their Consequences* (Sage 2014).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

Special-purpose agencies proliferated in the US from the 1920s onward, with the Port Authority of New York regarded by US historians as the 'mother' of all such 'special-purpose' or 'special district' governments. Sometimes the state (as distinct from the federal government) created agencies, such as urban development corporations, that used the vagaries of jurisdiction to muddy the waters of accountability, and often also of mere transparency. This development, which greatly undermined accountability in local governance, also took place in the UK, though flourishing only under the Thatcher regime.

When in 2017–19 I closely studied the failed 'smart city' project proposed by Google for a waterfront parcel of land in Toronto (together with my colleague and administrative law guru Alex Flynn), we found that the fact that the governance of most of the waterfront had long been downloaded by the three levels of government to the non-transparent agency Waterfront Toronto[14] really mattered. In that context I saw that 'the state' is not only fragmented through the workings of federalism and unclear jurisdictional dynamics but also has the opportunity and even the incentive of abdicating jurisdiction altogether, letting somewhat randomly chosen board members make decisions affecting a very large public.

Hence, it would be useful if scholarship were to investigate other periods and other fields of state action to see if the dynamic that Jen Raso has documented is visible elsewhere and in previous eras. In any case, it is healthy to have a professor of administrative law remind law and technology scholars that there is no such entity as 'the state', and that the proliferation of agencies is not just Weberian bureaucratic expansion gone mad but may act – very conveniently – to shield 'the state' from lawsuits enabled by the parallel proliferation of ombuds-type bodies and constitutional rights jurisprudence in the late twentieth century.

What Are Digital Data, Exactly?

I would first of all point out that just as the internet does not actually reside in the clouds, so too digital 'goods' are not ghosts hovering above

[14] Some readers may have seen international news coverage of this failed project. However, much of the international media coverage erred in stating that it was the city of Toronto that was Google's public sector partner. The project would likely have been rejected by the city early on if the city had been the 'partner'; but as it was the saga dragged on for years. The importance of the structure of Waterfront Toronto is not questioned in what has become the key account of the project, J O'Kane, *Sideways: The City That Google Could Not Buy* (Random House Canada 2022).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

the 'real' material world. Their financial and/or their use value might be mainly intangible; but they are embedded, inextricably, in vast material 'supply chains' that among other things use rare-earth minerals to produce smartphones and other emblems of digitality. These chains or networks may be primarily valued because of the digital data and digital predictions they enable; but they have to use electricity to run and to cool the humming computers in the server farms of Oregon and Washington State, where climes cooler than California's has meant a reduction in air conditioning expenses for the Silicon Valley giants that quietly buy land and build anonymous-looking server farms.[15]

There's a subsequent question often bracketed by law and technology scholarship: the question of what counts as 'data'. Many media studies and information scholars have argued that information can come in any format, from origin stories to oral histories to census returns: by contrast 'data' is a term best reserved for information that is suitable for quantitative and statistical manipulation and is collected and preserved digitally, in databases.[16]

And yet the emphasis on numbers may be somewhat misplaced. When my children were in elementary school they learned the rudiments of what was called, in their school, 'data science' (not math). A typical exercise had them asking classmates whether they preferred apples, oranges, or bananas, then producing simple bar graphs representing the aggregate information. This was a pen-and-paper exercise. Now, were their simple graphs examples of 'data'? Perhaps. But perhaps their graphs were not 'data', but simply information.

It does seem, however, that digital data, or perhaps better, datasets, are not merely internet-enabled versions of older practices of numeration (such as the mortality statistics collected by life insurance companies long before state pensions). As IP lawyers remind us, 'facts' are in the common law in the public domain. Facts (say, the French Revolution started in 1789) can of course be worked up into a dataset; and IP lawyers could then argue that the dataset is private property, since creating it involved some human creativity, not just facts in the public domain. What can be privatised is largely a political question (as seen in the historic case of the massive British 'enclosures' of land previously classified as 'commons').

[15] See Blum, *supra* note 7.

[16] A Blair et al (eds), *Information: A Historical Companion* (Princeton University Press 2021).

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

I suspect that it will be more useful for the kind of work represented in this volume to not worry about 'facts' versus 'data' but rather to distinguish 'data' in general (here including my children's hand-drawn bar graphs) from 'big data' – especially if we adopt Rob Kitchin's view that it is mobility and networking that identifies 'big' data. He gives the example of census returns as an example of a collection of data that is large in size but is not 'big data', since in many countries census data are by law kept separate and immobile. The immigration and border-control agencies in Jen Raso's case study could not have made use of census returns, for legal reasons.

While digital data is not ontologically unique, since contrary to the imaginary of IP law data always has a physical location or substrate, one curious feature of the digital world is that datasets do not exist on their own. Datasets are produced by some combination of humans and machines, and are managed, shared, and made to interact with other datasets in different ways. As is well known, the noted corporate success of the Alphabet/Google companies is largely due to that corporation's capacity to connect different datasets.

If we want to appreciate the mobility and interactivity of datasets it is useful to highlight those corporations known as 'platforms'. Platforms (Uber, Airbnb, Amazon) have infrastructural effects, Julie Cohen argues, very cogently. This means that they have effects on the physical world – as when road traffic changes as drivers frantically consult Waze to avoid traffic jams.[17]

That platforms are unique in their infrastructural effects is a plausible claim, though of course empirical studies are needed to confirm this insight. The infrastructural effects of Airbnb have been documented by many entities and municipalities (mainly under the rubric of the effect, which cannot be called unintended, of diminishing the local stock of long-term rental housing). Similarly, the Uber corporation has caused tens of thousands of additional cars to prowl city streets – and yet no regulatory system that I know includes congestion charges as part of the Uber licensing system.

The infrastructural effects of platform capitalism have not received as much attention as the actual or potential negative effects on privacy and

[17] The infrastructural effects of platforms may sometimes be self-defeating, however. In 2018 I was riding in a friend's car in Mexico City, and I noticed that constantly consulting Waze (a common behaviour in that car-clogged city) resulted in the alternative routes becoming choked with traffic. I concluded that too many drivers used Waze. In other words, Waze was the victim of its own success.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

the effects on existing economic relations (such as the fostering of a precarious labour force known as 'gig workers'.) But the vast literature on 'smart cities' contains many examples showing the effects on both municipal and privately owned infrastructure not so much of digital data but of how those data get used and circulate – effects that are not separate from the legal weakness of local governments in much of the common-law world.[18] It may be that data gathering is not the real issue, contrary to what civil-liberties organizations believe. It could be that *how* datasets are accessed, re-used and recycled across institutions and jurisdictions is the real problem, not so much data collection in itself.

A related point Julie Cohen makes is that just as the switch from older methods of governance and information gathering to digital data has infrastructural effects, there is also a parallel 'datafication' of infrastructure. The shared bicycles that the city of Toronto owns and that I use on occasion, for instance, are 'datafied' to the fullest extent: the app for the system shows me, as soon as I open it on my phone, a real-time map of the immediate neighbourhood that shows how many bikes are available in each of bike station and how many spots are open, if I want to leave a bike rather than take one. The bikes themselves likely have a location-data exhaust, presumably used to prevent theft and enable city workers to find 'lost' bikes. The bikes must also have a digital clock, since the first thirty minutes are free to anyone who like me has a yearly subscription.

I imagine every contributor to this book would agree that even if we stop ignoring the materiality of the so-called digital world, it is not a good idea to simply transpose categories and governance regimes developed for such infrastructural networks as roads and bridges to the realm of largely intangible digital goods and resources. Barbara Prainsack, a health data expert, suggested a few years ago that well-meaning arguments to the effect that many datasets should be considered legally as a 'digital commons' will founder if one simply uses existing categories of property and governance (state property; individual property; common property) and tries to apply these to digital datasets.[19]

One reason mentioned by Prainsack for the care needed to develop governance regimes for digital data is that the digital world includes 'open access' resources as well as resources (such as commercial secrets) that are not shared, including many that might be 'common' but not

[18] M Valverde and A Goel, 'Beyond the Legal Status of Cities: A Framework for Analyzing Multi-Level Governance' (2022) XIII(1) *Journal of Indian Law and Society* 146.

[19] B Prainsack, 'Logged Out: Ownership, Exclusion and Public Value in Digital Data and Information Commons' (2019) 6(1) *Big Data and Society* 1.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

common to all. Open access texts are qualitatively different from physical commodities in that they can be consumed in a 'non-rivalrous' manner, as economists say: many people can simultaneously read an online post, unlike conventional goods, which are subject to the zero-sum game of individual tangible property.

The mention of 'open access', a property regime increasingly relevant in the world of scholarly publications, may suggest to some that the jurisprudence and the political theory that has developed in regard to the physical (especially the environmental) 'commons' might be useful in elaborating policies for the extraction and the sharing of digital data. However, Prainsack argues forcefully that the physical commons cannot be transposed directly into the digital world. In her view, the much-touted idea of a 'digital commons' is not inherently impossible but would require much thought and discussion.

I raise Prainsack's point about the commons vs. open access for the scholars included in this anthology – and the readers – to consider. I would put the issue thus: given that digital data can be shared by many, even owned by many simultaneously, since contrary to Walter Benjamin in the online world the copies are the same as the 'original', scholars and public officials need to think about how the intangible components of digital data can be governed in such as way as to empower civil society in a progressive, equitable manner (instead of only empowering groups of people who all have a subscription to an online service or own goods with a shared element). The term 'the commons' is often thrown around by legal scholars with an anti-private enterprise bent, with little attention to the distinctions between state ownership, public ownership not controlled by the state, 'commons' and 'open access'.

To conclude: the two fascinating initial chapters provide an excellent introduction to many of the important issues surrounding 'data infrastructures', as well as raising important points about the infrastructural effects of some methods of data collection and use. If this comment serves to further the exciting collective discussion enacted in this book, it will have served its purpose.

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005

Downloaded from https://www.cambridge.org/core. UNIVERSIDAD AUTONOMA METROPOLITANA, on 05 Nov 2026 at 09:09:43, subject to the Cambridge Core terms of use, available at https://www.cambridge.org/core/terms. https://doi.org/10.1017/9781009425308.005