Science & Technology Studies

2/2016

Science & Technology Studies

ISSN 2243-4690

Subscriptions

Subscriptions and enquiries about back issues should be addressed to:

Jukka Syväterä
Pinni B 3002
33014 University of Tampere
Finland
email: jukka.syvatera@uta.fi

Subscription rates for 2014 are € 30 for individual subscribers and € 60 for institutions.
Rates include overseas postage.

Science & Technology Studies

Volume 29, Issue 2, 2016

Guest editorial

Articles

Discussion paper

Book reviews

Visit our web-site at

www.sciencetechnologystudies.org

Knowledge Infrastructures: Part II

Helena Karasti

Department of People and Technology, Roskilde University, Denmark / hkarasti@ruc.dk
Information Systems, Luleå University of Technology, Sweden / helena.karasti@ltu.se
INTERACT, University of Oulu, Finland / helena.karasti@oulu.fi

Florence Millerand

Department of Public and Social Communication, University of Quebec at Montreal, Canada / millerand.florence@uqam.ca

Christine M. Hine

Department of Sociology, University of Surrey, UK / c.hine@surrey.ac.uk

Geoffrey C. Bowker

Department of Informatics, University of Irvine, CA, USA / gbowker@uci.edu

The papers presented here were submitted in response to a call for papers that sought to draw together the current state of understanding of knowledge infrastructures from the viewpoint of STS and to provide a basis from which to evaluate the distinctive contribution that the theoretical resources of STS were making within this territory. That call for papers produced a high level of response, providing a clear indication that STS scholars are indeed taking knowledge infrastructures seriously, and that the study of infrastructures is providing fruitful ground for developing insights into STS's core concerns with interrogating the complex, emergent sociotechnical systems that pervade the contemporary world. The initial call for papers produced more successful submissions than could be accommodated in a single issue of the journal, and hence the envisaged special issue will, in fact, extend across multiple issues of which this is the second.

In the previous issue of *Science & Technology Studies*, we presented an initial batch of three substantively very different studies: Wyatt et al. (2016) explored the treatment of controversy within the production of the Wikipedia entry relating to schizophrenia genetics; Parmiggiani and Monteiro (2016) examined the production of infrastructures relating to the monitoring of environmental risk in offshore oil and gas operations; and Boyce (2016) analysed the work of connecting infrastructures for public health surveillance. Despite the differing substantive foci we were able to draw out some significant cross-cutting themes. The issue of scale received considerable attention, as the papers each explored what were on the face of it large scale infrastructures but were sustained by contingent connections forged across macro-level visions of possible outcomes and diverse forms of micro-level work developing technologies, connecting systems, generating content, overcoming obstacles and managing breakdowns. Our editorial (Karasti et al., 2016) took a reflexive turn, considering the significance of the methodological choices that underpinned these studies of infrastructures and the intransigence of some aspects of infrastructure in the face of our attempts to comprehend them. We noted that the choice of where and how to study such infrastructures involves some significant decisions on the part of the analyst in terms of the focus and level of examination (Larkin, 2013) and also the individual sites and relations to study when a large scale infrastructure can appear at first sight to be everywhere at once and yet nowhere in particular. While the choice to adopt the infrastructural inversion (Bowker, 1994) positions the infrastructure in the foreground and focuses attention on the many forms of work that bring into being and sustain the infrastructure, this initial methodological choice leaves many others for the analyst to navigate.

A further theme that resonated through the articles was the issue of invisibility, whether that concerns the taken-for-granted nature of the infrastructures themselves or the habitual lack of attention in many public spaces to the various forms of work that sustain them. Invisibility has been a fundamental concept within STS studies of infrastructure (e.g. Star & Ruhleder, 1996; Star, 1999; Bowker & Star, 1999) and within this batch of papers the notion of invisible work was clearly apparent, and yet across the three papers invisibility played out in quite different ways for both actors in the setting and analysts. Issues of tension, friction and repair also recurred across all three papers, as did the management of ambiguity and uncertainty. Actors and STS analysts sometimes shared a concern with how far to tolerate ambiguity and where to strive for a more concrete solution. Specific relations of accountability determine what counts as a "good enough" knowledge infrastructure for purpose and underpin the accounts that both actors within the setting and their STS guests offer up.

These emergent themes of scale, invisibility, tension, uncertainty and accountability continue to resonate across the four pieces presented in this second instalment of the special issue on knowledge infrastructures. In the rest of this editorial we will introduce the pieces and then draw together, briefly at this point, additional themes that emerge at this stage. In a future editorial we will step back to review these themes across the full collection of papers in order to evaluate the current state and emerging challenges for STS studies of knowledge infrastructure as represented here.

Articles in This Second Part of the Special Issue

Three articles and one discussion piece are presented in this second part of the special issue. The special issue opens with an article by Masato Fukushima on value oscillation in knowledge infrastructures. By 'value oscillation', Fukushima is referring to the constant to and fro in knowledge infrastructures in the making between participants being told high of the potential positive value of the infrastructural work (for the good it will do in the world) and being warned of its potential negative value (for the harm it can do to one's career to perform service work). The oscillation refers to the constant tacking back and forth between the two. He explores this in two case studies – one of an open database and data library of natural products, and the other of a database used in a drug discovery pipeline. Wrapping these rich empirical analyses is a theoretical argument about linkages that science studies scholars might make with earlier work (notably Marx, Godelier and Lévi-Strauss) through recognizing the resonance with their uses of versions of infrastructure and superstructure. He argues that in a sense we have to our detriment lost touch with our own invisible intellectual infrastructure.

The concept of value oscillation is a particularly good one for understanding knowledge infrastructures in general. From the science studies tradition, particularly actor-network theory, there has been a tendency to see people as either translating the interests of others or having their own interests translated – so that ultimately the black box that emerges is unary and univocal. Fukushima's analysis suggests that at different moments one can switch between different value systems without necessarily realizing the contradiction – the value, one might say, inheres to the specific situation, not to a single actant. This move opens the possibility for new understandings of the distribution of moral qualities in dense networks of humans and non-humans.

The second paper "Building knowledge infrastructures for empowerment: A study of grassroots water monitoring networks in the Marcellus Shale" focuses on the issues of power and empowerment in the building of knowledge infrastructures for citizen science. Kirk Jalbert studies nongovernmental environmental monitoring networks engaged in water monitoring in a Northeast U.S. area where oil and gas are drilled using hydraulic fracturing, a controversial method of extraction. Jalbert reasons that the lack of transparency in the poorly regulated practice of hydraulic fracturing has made it a particularly germane domain for civil society sector involvement. Citizens become active in attempts to understand the environmental impacts of the oil and gas business in their own backyards.

Jalbert has studied longitudinally two grassroots environmental monitoring networks of citizens. One of them is a coalition of advocacy groups and the second is a large network managed by academic institutions. The networks, concerned for public heath and environmental risks introduced by shale oil and gas extraction, assemble resources for monitoring, collect data and build alliances. They, according to Jalbert's argument, construct distinct knowledge infrastructures that can empower participants to question scientific assessments made by more powerful institutions, participate in public debates and influence regulatory decision-making.

With focus on a discourse of power and empowerment, Jalbert's paper offers a theoretical contribution to facilitate understanding of the conditions under which marginalized stakeholder groups take part in shaping knowledge work and building knowledge infrastructures in order to address complex scientific and environmental issues. Aligning with current understanding of knowledge infrastructures as emerging and adaptable, Jalbert finds that while the formation of knowledge infrastructures can reproduce established relations of power, the grassroots groups are able also to tactically alter power dynamics and redistribute resources to their advantage. This is an encouraging finding for the participation of and influence by marginalized stakeholder groups in the face of the continuing struggles involved in dealing with environmental problems and associated policy struggles as laid out in the conclusions of the paper.

The third paper "Making knowledge in boundary infrastructures: Inside and beyond a database for rare diseases" investigates the ways in which infrastructural issues come to matter in the production of knowledge in the social worlds of rare diseases. Eric Dagiral and Ashveen Peerbaye conducted a four-year ethnography of the "Rare Diseases Platform", a European-level entity created in the early 2000s and located in Paris (France). They analyzed in detail a relational database devoted to rare diseases and orphan drugs that represented one of the major achievements of the large and complex network of individuals, institutions, and practices that the European Platform created.

Their study takes up the concept of "boundary infrastructure" and explores its practical and theoretical implications, by examining how a wide array of actors negotiate the place and forms of knowledge production in relation to many of the other goals they pursue. Indeed, in contrast to situations in which knowledge production is the core legitimate focus of the collective action (e.g. in laboratories or scientific collaboration networks), the involvement of actors and communities around the database for rare diseases extends well beyond this purpose, so that, knowledge production, as one of many outputs of infrastructural work, needs to be articulated with other matters of concern, some with explicit political and moral aspects.

Dagiral and Peerbaye's contribution suggests two main claims. One is the political nature of the distinction between knowledge and 'mere' information, as this demarcation line may embed competing visions on what the infrastructure should be and what it should do in relation to the collectives involved (e.g. researchers, patients, the general public, institutions). In looking at the ways in which what counted *as* knowledge infrastructure and what counted *in* a knowledge infrastructure was materially enacted within the database, the authors found that the category 'information' rather than 'knowledge' became the category of choice, under which participants in the Platform could frame themselves as involved in the "fight against ignorance of rare diseases", as much as in the production of "novel biomedical knowledge". A second claim recognizes that "infrastructural inversion", this STS methodological lens for the analyst to scrutinize all the activities that warrant the functioning of an infrastructure rather than those that it invisibly supports, may also be constitutive of the practices of the actors themselves. In this case practicing infrastructural inversion served the communities involved to articulate knowledge production with other forms of mobilization, as they negotiated the political, moral and epistemic dimensions of the boundary infrastructure they contribute to. In doing so, the database became reshaped and reconfigured, for instance through classification activities (e.g. choice or deletion of heading for diseases names that didn't "sound right" or that seemed too "complicated" for the patients), thus presenting itself in a state of continued reconfiguration.

In the discussion piece, Kalpana Shankar, Kristin Eschenfelder and Greg Downey use the lens of knowledge infrastructures to shed new light on some well-established practices in their discussion paper "Studying the History of Social Science Data Archives as Knowledge Infrastructure". Social science data archives have been in existence for decades and yet, the authors argue, their role in the development of social science disciplines has been little acknowledged. They suggest that there has been minimal critical attention to the precise nature of the unfolding relationships that constitute social science data archives as infrastructures and in turn shape the possible future directions of the disciplines. Intriguingly, social science data archives pre-date the current era of open access and digital data and provide, the authors argue, for some interesting comparisons with contemporary cyberinfrastructures. Shankar et al. observe early shifts towards data intensive forms of work in social science disciplines that prompt intriguing comparison with contemporary developments. Some interesting international dynamics also emerge, as social science data archives are developed on both sides of the Atlantic and fit themselves into the distinctive arrangements of professional organizations, governmental expectations and funding prospects within each context.

Focusing particularly on quantitative social science data archives, Shankar et al. describe a complex ongoing evolution and mutual shaping of archives and fields of knowledge production, with shifting rationales and sets of relations and an ongoing struggle to justify the labour needed maintain an archive in the face of competing pressures. They suggest that dealing with rupture, discontinuity and breakdown is inherent in the work of infrastructuring, as much as building, creating and forming relationships. Studying the history of social science data archives through the conceptual apparatus offered by STS approaches to infrastructuring provides, the authors suggest, an interesting case to compare with contemporary efforts in other disciplines when considering what makes for a sustainable knowledge infrastructure.

Reflections and Emerging Themes

Three major commonalities emerge across this rich set of material. The first of these is methodological: each in its own way is performing an act of infrastructural inversion. The authors are looking at what happens when you focus your attention on the infrastructure itself, acknowledging that it has a history and a context and that it takes work to bring it to life. The second common theme flows from the methodological commitment. In their different ways, these articles demonstrate how knowledge infrastructures are performative of the knowledge being produced – they are not passive backdrops. The work of Shankar et al. on the social science archive is interesting for this reason as your theories depend on the kind of archive you can build. Thus for the longest time, ecology as a discipline was tied to its archive of one meter squared plots of land or memory studies to its archive of laboratory results (preventing, in the latter case, the development of social theories of memory). Or, in the case of Dagiral and Peerbaye, the 'background' database conjures knowledge into particular forms. The third theme is a development of this perspective: both in their shaping and deployment, knowledge infrastructures are core sites of political action – from the need to represent and acknowledge invisible work to the need to build infrastructures which are sensitive to multiple perspectives. In addition to the concerns with scale, invisibility, tension, uncertainty and accountability identified within the first batch of articles, this issue focuses our attention particularly on concerns with power, marginalization and values. Fukushima highlights the shifting territory of values in relation to infrastructural work and outlines a set of theoretical resources that could bring to the fore a new sensitivity and nuance to the notion of infrastructures as a site of power. Jalbert focuses on discourses of empowerment and the struggles over the potential for marginalization that pervade citizen involvement in infrastructures enabling grassroots environmental monitoring. Dagiral and Peerbaye follow the articulation of infrastructural work with matters of political and moral concern, finding that the distinction between knowledge and information can be highly charged and consequential within struggles to meet the needs of the various collectives implicated in the development of databases depicting rare diseases.

Across these three articles, then, we encounter struggles over power, values and voice at the very heart of infrastructural work. Such concerns are less immediately apparent in the discussion paper from Shankar et al., but are nonetheless present. In setting an agenda for STS-inflected study of social science data archives the authors make clear that these archives too act as sites for negotiation of power, voice and values. Social science data archives, for Shankar et al., become sites where competing versions of the value of different forms of labour and knowledge production collide, where a political will to perform particular kinds of governance and foster certain institutional arrangements is enacted and where visions to move whole academic disciplines towards an envisioned data intensive future play out.

Bibliography

Bowker GC (1994) *Science on the run: Information management and industrial geophysics at Schlumberger, 1920–1940*. Cambridge: MIT Press.

Bowker GC & Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge: MIT Press.

Boyce AM (2016) Outbreaks and the management of 'second-order friction': Repurposing materials and data from the health care and food systems for public health surveillance. *Science & Technology Studies* 29(1): 52-69.

Karasti H, Millerand, F Hine CM & Bowker GC (2016) Knowledge infrastructures: Part I. *Science & Technology Studies* 29(1): 2-12.

Larkin B (2013) The politics and poetics of infrastructure. *Annual Review of Anthropology* 42: 327-343.

Parmiggiani E & Monteiro E (2016) A measure of 'environmental happiness': Infrastructuring environmental risk in oil and gas off shore operations. *Science & Technology Studies* 29(1): 30-51.

Star SL (1999) The ethnography of infrastructure. *American Behavioral Scientist* 43(3): 377-391.

Star SL & Bowker GC (2002) How to infrastructure? In: Lievrouw L A & Livingstone S L (eds) The handbook of new media. *Social shaping and consequences of ICTs.* London: Sage Publications, 151-162.

Star SL & Ruhleder K (1996) Steps toward an ecology of infrastructure: Borderlands of design and access for large information spaces. *Information Systems Research* 7(1): 111-134.

Wyatt S, Harris A & Kelly SE (2016) Controversy goes online: Schizophrenia genetics on Wikipedia. *Science & Technology Studies* 29(1): 13-29.

Value Oscillation in Knowledge Infrastructure: Observing its Dynamic in Japan's Drug Discovery Pipeline

Masato Fukushima

The School of Arts and Sciences, The University of Tokyo, Japan / maxiomjp@yahoo.co.jp

Abstract

This paper analyses the dynamics of assigned values in two cases relating to the knowledge infrastructure of the national programme in Japan that develops drug discovery: in establishing a database of natural product compounds and in constructing a library of virtual compounds. The concepts of value oscillation and of the M-B (Marx-Bowker) index are proposed to designate the fluctuating appreciation of infrastructure value by its builders. These concepts combine insights from classical Marxist thought on the infrastructure/superstructure distinction (neglected in recent studies on infrastructure in STS) and Bowker's infrastructural inversion. Though value oscillation is almost ubiquitous in the development of any infrastructure, in the cases considered here, it takes peculiar forms because of the complex interaction of the material and knowledge infrastructures. It is widely distributed in the sub-layers that support the autonomy of these knowledge infrastructures and is a precondition for knowledge infrastructures to function as delineated entities.

Keywords: value, infrastructure, drug discovery, oscillation, structure

Introduction

Concepts are the ways through which we see the world, and scholars have long realized that apparent conceptual lucidity may hide winding paths that can produce a variety of contradictory nuances, leading to persisting controversies. Thus, academics from various fields have traced the meandering former paths – including their etymology – of such concepts as subject/object (Williams, 1976), liberalism (Hayek, 1982), society (Luhmann, 1980), and even 'thing' (Heidegger, 1968; Latour, 2005; cf. Fukushima, 2005).

From this perspective, the recent rise of so-called 'infrastructure studies' in STS apparently falls short of reflecting infrastructure's conceptual genealogy, while there are numerous concrete examples analysing what infrastructures are and have been. A brief review of foundational works indicates that infrastructure is usually conceived as a collection of such conventional prototypes as roads, water conduits, and electricity; later, infrastructure took on new, extended meaning in terms of such phenomena as communication, information, and even knowledge (Star & Ruhleder, 1996; Star & Bowker, 2002; Edwards et al., 2007; Bowker et al., 2010). This approach to defining the subject, however, is plagued with historical amnesia in terms of its intellectual genealogy. The concept of historical amnesia, which will become clearer through the rest of this article, can briefly be defined as blocking recourse to the proper legacy of the past for systematic reasons, often organizational, political, or even social (cf. Bowker & Star, 1999, ch. 8), in STS scholarship.

By way of demonstration, in 1978, *Current Anthropology* published an article by Maurice Godelier and colleagues titled 'Infrastructures, Societies, and History' (Godelier et al., 1978). Godelier is well known for his innovative endeavours to unify Marxist anthropology and French structuralism (cf. Godelier, 2011). His paper was intended to redefine the Marxist version of infrastructure to solve the evident contradiction between societies in which such elements as kinship or religious institutions seem to dominate *contra* Marxists' classical tenet that modes of production determine other societal factors. At stake here is the assumption of historical materialism – namely, that the base, or 'infrastructure', of a given society is defined as the productive forces and social relations of production that unilaterally determine the rest of society, the superstructure (henceforth capitalized to represent their unitary character; Marx, 1973). Large amounts of energy have been expended to improve or even alter this rather rigid framework (Lichtheim, 1971; Howard & Klare, 1972; McLellan, 1979), and Godelier and colleagues radically expanded the Marxist understanding of Infrastructure to what he calls *idéel* reality, a notion inspired both by phenomenology and structuralism, consisting of thought and language, knowledge of nature and tool usage, and taxonomy and classification (Godelier et al., 1978: 764).

The concept of Infrastructure has been important since Marx (1973) formulated it in his *Grundrisse*[1], and it has underpinned the social sciences to various extents. However, the canonical collections of infrastructure studies noted above seem to be silent about this specific line of Marx's intellectual legacy (cf. Carse, 2012: 542–44). The reason for this amnesia may be that, aside from the fact that the term 'base' is more often used in Marxist terminology, the preferred theoretical approaches of these authors of foundational studies, such as symbolic interactionism (Star & Ruhleder, 1996; Star, 1999), system theory *a la* Parsons and von Bertalanffy (Hughes, 1983), and ANT/SCOT (others), have drawn researchers away from this tradition. In fact, the merit of these approaches is undeniable: by 'clearing and erasing' the past traces (Bowker & Star, 1999: 257), STS has produced a fairly large number of empirical, fine-tuned studies of more specific technical infrastructures.

While admitting the advantages, I claim that at least three major problems have been overlooked by not critically conversing about and confronting the Marxist legacy: 1) What is Infrastructure? 2) How does it work? 3) How do we understand it?

1) In STS, the question of what infrastructure is usually relies on Star and Ruhleder's (1996) foundational definition, which presents eight features that are seemingly distinct from the Marxist concern with the mode of material production as the unitary basis for Infrastructure. Overlooked here is not so much the chance of comparing the two as a missed opportunity to refer to the extremely rich inventory of efforts to revise the latter after its initial formulation in *Grundrisse*. In addition to the ensuing attention to the pivotal influence of the Superstructure on its counterpart (McLellan, 1979; Anderson, 1976), to be discussed below, efforts have also made to find alternatives to modes of production, such as consumption (Bataille, 1988; Baudrillard, 1981) or exchange (Karatani, 2014; cf. Polanyi, 1944; Sahlins, 1972). Among these, the most striking case is the above-mentioned innovation of Godelier and others in adding structuralistic elements like taxonomy and classification into Infrastructure, directly leading to Bowker and Star's (1999) similar claim 20 years later, a foundational case for the present treatment of knowledge infrastructure. By failing to examine such preceding efforts, STS researchers have clearly missed the chance of incorporating certain insights into, for example, how the *idéel* system works together with a more classic mode of production, which could provide clues on the link between the knowledge infrastructure and the wider economy in the present argument[2].

2) The way Infrastructure works is tightly related to its counterpart, Superstructure; for Marx, his followers, and his critics, however, the relation between these two poles has been the

target of extensive examination and polemics because critics regard the Marxist definition of Superstructure as too loose, as it includes virtually everything except modes of production. In contrast, STS circles seem to have shown relatively ambiguous attitudes to questions concerning the effect of infrastructure and how it is constructed. Again, the point here is not to adopt the unilateral determination of the earlier formulation, but to re-examine ensuing efforts to reformulate the very meaning of determination: for instance, some even argue that Althusser and Balibar's (1970) concept of overdetermination is essentially in line with the notion of complexity (Shiozawa, 2002), possibly contributing to the present discussion of how multiple infrastructures interact with each other (cf. Vertesi, 2014).

This determination thesis is also directly related to the problem of power in terms of the class that dominates Infrastructure. This element, as part of the legitimate vocabulary of political sociology, seems to have some shadowy resonance in contemporary infrastructure studies; however, references to the issue are both hesitant and lacking theoretical cultivation in terms of what kind of power is related to the issue (cf. Edwards et al., 2007: 24–31; Edwards et al., 2009: 371).

3) The question of the understanding of Infrastructure and the value directly attached to such an act of recognition, the central theme of this article, requires full exploration. In the Marxist conceptualisation of these paired concepts, the latter represents the surface and visible values that apparently dominate society, whereas the former is submerged. The Marxist strategy of historical materialism, in essence, is to destroy this naive view of the dominance of such surface values by 'turning Hegel on his head' (Marx, 1976/81), an act of inversion in the face of the ladder of values that exists both in Hegelian idealism and in the real world, highlighting this apparent inferiority of the value of Infrastructure in order to reveal its determinant power.

Thus formulated, the following arguments concerning infrastructure in STS have followed a similar path without attending to its intellectual ancestry. References to the negative evaluation of infrastructure have been scattered in the preceding body of research, in which infrastructure is described as boring and unexciting (Star, 1999: 377), as maintained by undervalued and invisible workers (Star & Bowker, 2002; Bowker et al., 2010: 98), and as often characterized by 'tension' with regard to its value (Edwards et al., 2013: 26)[3]. One description of the ambivalent aspect of treating taxonomy as infrastructure summarizes the issue here:

> "Being treated as infrastructure has hitherto been a dubious honour. While being considered essential gives one a certain amount of leverage, it also means one risks being taken for granted and neglected in the face of other, more prominent topics." (Hine, 2008)

When this idea is extended to metadata, things do not seem to be radically different:

> "All recognize metadata's potential value, but when the rubber meets the road, an unfunded mandate to be altruistic [...] does not prove highly attractive." (Edwards et al., 2012)

Thus, although the conflict of values regarding infrastructure has been a matter of constant, if sporadic, concern even within the study of infrastructure in STS, its formulation lacks the consistency of the Marxist tradition in dealing with their own version of Infrastructure.

Value Oscillation and the M-B (Marx-Bowker) Index

In this article, drawing upon the last point above, I focus on this discrepancy: whereas the power of infrastructure is recognized, the practices related to it, such as service to others and its maintenance and repair, are not highly ranked in the existing value system, being often regarded as invisible and even 'boring'. Because of this particular nature of infrastructure, Bowker (1994) proposed 'infrastructural inversion', a strategic analytical operation to bring hidden infrastructure to the surface and expose its importance. My claim in this paper is that this particular operation is, in essence, structurally isomorphic with the Marxist operation of 'turning Hegel on his head', different only in terms of its focus and scope of theorization.

These operations have thus far been confined to analysts' strategies, whereas my examination relates to how practitioners *in situ* regard the value of infrastructure and its related practices. Conflict, contradiction, and oscillation (as is evident to some extent in the preceding monographs on the issue) are expected from this approach, precisely because practitioners are the ones who develop and maintain infrastructures. Hence, I will adopt the term 'value oscillation' for describing this aspect of fluctuating attitudes, between these two poles of the recognition of its supportive value and avoidance of its shadowy character.

This observation relates intrinsically to the very concept of infrastructure itself, which is almost oxymoronic: though infrastructure exerts immense power as it structures other things, it is inferior (*inferus*, inferior from *infra*, in Latin) because it lacks surface value. By way of analogy, Weick and Westley (1999) have claimed that 'organisational learning' is an oxymoron because organizing is a process of reducing complexity, whereas learning increases it; hence, organisational learning is rare. Infrastructure as oxymoron exists rather steadily but generates a double-bind (Bateson, 1972) for its concerned practitioners, owing to its intrinsically opposing vectors of value. Because its value oscillates between these two poles, like other double-bind situations, it is hung in indeterminacy. Hence, my term, 'value oscillation', is more adequate than conventional expressions like 'value conflict' or 'contradiction': these are both too macroscopic, and they also easily connote a 'dialectical' solution of cancelling the contradiction out (*aufheben!*), which, I believe, rarely happens in a double-bind.

To describe this zig-zag movement of indeterminacy, I introduce a second term, the 'M-B (Marx–Bowker) index', to show the degree of appreciation for the invisible infrastructure values. Here, 'infrastructure' is defined not only as the material entity designated by the term, but also the wider assemblage of activities involving quasi-public services to others, works of a sub-contractive nature, and backstage efforts including those indirectly related to infrastructure building. The juxtaposition of these two names signifies the intrinsic continuity of the two approaches in terms of inverting the underlying value, while simultaneously emphasising the practitioners' view and action; in addition, the index becomes an easily visualized means for observing the oscillating attitude of the concerned practitioners.

Presupposing practitioners' recognition of the validity of any given infrastructure, the M-B index is defined as concerned practitioners' observed degree of commitment to developing and maintaining a given infrastructure: hence, a high index means a high degree of commitment to it, and a low index implies avoidance of such commitment. In this paper, this index is used as a figurative tool for visualising the observed oscillation of practitioners' attitudes as demonstrated by both their discourse and their actions vis-à-vis the issue of building and maintaining infrastructure. One may argue that such values behind our actions are too complex to be adequately identified with this index. This argument admittedly has some truth; however, I claim that when we focus sharply and directly on the issue of building and maintaining infrastructure, we may reduce it to a simple question of whether one promotes it or avoids it, though there may be intermediate choices with accompanying complementary reasons. Such focused action and discourse, along with any oscillation, are in fact observable, reaffirming Geertz's (1973) classical formulation of cultural practice as a public vehicle of meaning.

My own research is based upon ethnographic observation, and I use such relative expressions as 'high' and 'low' with regard to the M-B index. However, my approach does not preclude the possibility of substantiating the claims by using questionnaires, though I did not attempt such in this project. In such a case, the M-B index could be tentatively quantified, with zero meaning the practitioners' avoidance of any commitment to infrastructure building, and 5 or 10 showing full commitment to its construction, thus expressing a continuum.

Some laboratory studies seemingly present cases of apparent value oscillation concerning the ambivalent role of research tools and related work practices (Clark & Fujimura, 1996; Gaudilliere & Löwy, 1998; Joerges & Shinn, 2001; Mody, 2011), the interchangeability of epistemic things and research technology (Rheinberger, 1997; Joerges & Shinn, 2001), and the problem of the fluctuating status of such tools in the hierarchy of credibility in laboratory settings (Clark & Fujimura 1992: 16)[4]. However, infrastructure goes far beyond the limited scopes of laboratory and disciplinary boundaries, and its multi-layered character increases the complexity of analysing value oscillation, as it is distributed across diverse spaces and various layers. It is further complicated when extended to its knowledge aspect, wherein 'infrastructure' generally signifies the whole set of heterogeneous elements of databases, computerization, grid systems, e-science, and so on without (thus far) a proper definition (Edwards et al., 2009; Edwards et al., 2013).

In fact, the question of how value oscillation takes shape arises in light of the ongoing momentum and extensive influence of computers, information, and even data science as 'science' (Hine, 2006b, 2008; Edwards et al., 2007; Edwards et al., 2009; Bowker et al., 2010; Edwards et al., 2013). We can presume that these factors push the M-B index both upwards – because the halo of new science attracts devotion – and downwards because 'infrastructuralization' is avoided in such areas compared to established engineering efforts to maintain roads, water, and electricity. Thus, possible strategies that concern the dynamics of knowledge infrastructure become a question of concern.

The Research Subject

The remainder of this article will discuss 'value oscillation' as it relates to the knowledge infrastructure in two distinct case studies, both related to attempts to build a sort of database as part of the larger scheme of Japan's national policy of developing an infrastructure for drug discovery (*sôyaku-kiban*) that is academia-based. First, we look at a faltering endeavour to establish a database of natural product compounds to make the search for drug seeds more effective and to facilitate basic research for ligand–protein interaction. Second, we will examine the ongoing construction of a large-scale virtual library of chemical compounds, using a world-class supercomputer.

The analytical focus in these case studies is twofold. The first is on how value oscillation is observable in the multi-layered infrastructures wherein these databases are embedded. The schemes for building such drug discovery infrastructure require coordinating and constructing various layers of sub-infrastructures simultaneously, providing intriguing examples of how the problem of value oscillation is approached in the different layers beyond the confines of the specific databases that are the main focus.

The second focus is how this issue is related to the context of knowledge and material interaction. As drugs are material entities that require a vast amount of heterogeneous knowledge, the development of the knowledge infrastructure in this context is closely related with its material counterpart in producing drugs. By highlighting these two aspects, this paper examines the various appearances of value oscillations throughout the complex, multi-layered character of the knowledge infrastructure and how the practitioners deal with the situation in each contextual effort, along with the consequences[5].

Background: Drug Discovery Infrastructure as Knowledge Infrastructure

Drug discovery is a hugely complex process that demands a vast amount of heterogeneous knowledge and related infrastructure, beginning with finding the proper target proteins and drug seeds and progressing to a range of steps from animal testing to clinical trials, which include Phases I to III (Epstein, 2007; Petryna, 2009; Keating & Cambrosio, 2003, 2012). Behind the policy idea of developing a national drug discovery infrastructure lies the fact that the productivity of drug discovery has decreased sharply despite the growing knowledge and technology related to the process, and controversies have occurred about its possible causes (Epstein, 2006; Ryzewski, 2008; Bartafai & Lees, 2006; Kubinyi, 2003). Drug companies have thus urged governments to promote the idea of outsourcing such development to academia, which is supposed to be able to bear greater risks. This idea gained momentum when the National Institutes of Health (NIH) in the United States published their Roadmap Initiative for Biomedical Research in 2003 to promote constructing an academic drug discovery infrastructure in the form of public chemical libraries and screening centres for academic use (Wikstrom, 2007; cf. Mazzucato, 2013). In response, the Japanese government launched their version of the policy (Fukushima, 2015).

Here, I comment briefly on the peculiarities of considering the drug discovery infrastructure as a particular type of (knowledge) infrastructure. Despite the general usage of the term *kiban* in policy discourse, 'infrastructure' here means the very specific purpose of producing drugs, as opposed to more general infrastructures like roads and the Internet. The alternative term, the drug discovery 'pipeline', also connotes the horizontal integration of the temporal procedures from bench to bedside. Thus, some researchers prefer using terms like 'platform' (Keating & Cambrosio, 2003) to highlight the horizontal assemblage of knowledge and material rather than the term 'infrastructure'.

Nevertheless, the term 'drug discovery infrastructure' has its own legitimacy. First, this process consists of multi-layered entities, ranging from the national plan to specific institutions to the laboratory level, where various aspects of infrastructure-like characteristics can be spotted, exhibiting similarity with other types of more conventional infrastructures, such as databases open to academic purposes (Star & Ruhleder, 1996; Edwards et al., 2007; Edwards et al., 2013).

Second, the process includes a unique entanglement of materiality and knowledge. Although drugs are an industrial material, they are also what Barry (2005, borrowing from Bensaude-Vencent & Stengers, 1996) calls 'informed material', which requires a huge complex of knowledge from protein science, chemistry, and medicine, wherein the elements of the knowledge infrastructure play pivotal roles.

In the following sections, the sub-institution levels are given priority for the detailed analysis, but higher levels are by no means unrelated. The focal institution is RIKEN, a public research institute representative of basic science in Japan. RIKEN's involvement in the infrastructure plan is the main background. RIKEN has experienced a series of ups and downs, from its pre-war status as the pivotal nexus of research and industry to post-war decline and revival in recent years in the form of national genomic and postgenomic research projects given by its supervising ministry (RIKEN, 2005). After a series of major science projects, such as Protein 3000 Project (Fukushima, 2016), RIKEN launched a plan to establish a more tightly woven infrastructure for drug discovery with a more effective organization of its branches, which had not previously been tightly combined with one another. The following cases both fall within the larger scheme of RIKEN's policy[6].

The Chemical Biology Centre as Future Quasi-Infrastructure

The first major topic of this paper is the faltering effort to establish NPEdia (whose name is abridged from 'Natural Products Encyclopaedia'), a database for public use. This project was implemented along with the development of NPDepo, a public library of natural product compounds. These plans were launched in parallel with a government plan to establish a national library of chemical compounds open to academic use. Natural products are the chemical entities produced by living things, such as microbes, plants, and marine organisms. These entities have bioactivity—namely, the effects exerted on other living things. This genre of research has had extensive relations with drug discovery, most notably in the case of antibiotics extracted from fungus, such as penicillin, or recent searches for plants and marine organisms to provide new seeds for drugs (Fukushima, 2015).

This specific idea was promoted by a number of RIKEN's leading laboratories; among them, the antibiotic laboratory (of more than 60 members), which boasts a long genealogy of preceding laboratories in the same genre of research (Ueno, 2008), has taken the pivotal role. This infrastructural innovation was accompanied by an organizational plan to establish a new centre for an emerging hybrid science called 'chemical biology', wherein chemical compounds are used to regulate and probe the activity of life phenomena. In the United States, the above-mentioned Roadmap Initiative policy to promote chemical biology includes a public chemical library and screening centre (Wikstrom, 2007); however, controversies have developed between the NIH and leading scientists there over how best to orient chemical biology for drug discovery (Fukushima, 2013).

Like Matryoshka (Russian nesting) dolls, the problem of value oscillation vis-à-vis the development of various layers of infrastructure can be observed on various levels, from RIKEN itself down to the laboratory practices. One of the focuses here is the centre level, within which the NPEdia/NPDepo complex is situated. The chemical biology centre plan was once intended to establish one of the hubs for the coming drug discovery infrastructure, both within and without RIKEN, symbolising RIKEN's commitment to connect academia and industry by implementing governmental biomedical policy more directly. For that purpose, RIKEN has increased the number of time-limit centres to improve the infrastructural functioning of various kinds of large facilities, libraries, and databases (RIKEN, 2005). The chemical biology centre, under the leadership of the antibiotic laboratory mentioned above, was once part of this long-term plan. The centre was intended to be equipped with not only the database and library, but also with various assay systems as well as high-throughput machinery to enable the rapid examination of ligand–protein interactions, for public service as well as for their own research. However, a contradiction has thus been revealed about what the centre was meant to be from the beginning.

First, despite the official emphasis on the infrastructural objective, the promoters of this plan also intended to use the centre to pursue their own research innovation. I observed this divided intention during my visit to the laboratory, where a large part of the researchers' energy was spent preparing for the coming centre. In fact, the main members of the laboratory were subdivided into a number of teams, each consisting of a team leader and several members and technicians, each tasked with various infrastructural obligations, such as improving the high-throughput machinery, establishing new assay systems, and collecting and classifying materials for NPEdia/NPDepo.

Here, the division of labour is not confined to scientist/technician distinctions; almost all the scientists were also assigned to one or more infrastructure-related tasks for the coming centre. However, the distribution of such workloads was uneven, with some teams doing basic work like collecting materials and organizing information, while others were only doing their own homework[7].

Many examples of value oscillation occurred in this complex distributions of workloads. For instance, the 2008 intra-RIKEN official report, which is issued every seven years about the activity of the laboratory, was concerned mainly with the activities directly related to the infrastructural aspects of the future centre, while the outcomes of the individual research activities were given only passing references. Thus, its M-B index for emphasis on infrastructure elements was high. However, these individual papers were published in major journals and reported in a separate annual record about the laboratory's academic activity (interview, 24/6/2008). In addition, the uneven distribution of the infrastructural workloads led to some rather cutting remarks by some of the staff about their colleagues' work. For instance, after the official interview, one of the team leaders suggested to me that there would be no need to interview some of them, as their work was nothing but technicians' work, not that done by scientists. Such remarks clearly demonstrate a contrastively low M-B index score despite the laboratory's general policy (field note, 20/11/2007).

The changing discourse of the laboratory head was a living example of such value oscillation. In a meeting with the whole laboratory, for instance, he rather openly warned those who were then too keen to do service work for outsiders about their infrastructural duties, such as examining the bioactivity of the compounds entrusted to them. Even though these are the sort of preliminary duties that the future centre would be expected to carry out, he underlined the possible danger of doing too much service work for outsiders, which could decrease the quality of their own scientific activity (field note, 15/4/2008). On another occasion, he suggested that the staff collaborate on their entrusted work if they found the job interesting enough to do as part of their own research (field note, 13/11/ 2007). This impressive degree of ambivalence was observed throughout the laboratory, even, as demonstrated, with the same leader, with the M-B index score seeming to change daily, like stock prices.

NPEdia and NPDepo: Between Material and Knowledge Infrastructure

The above example, wherein the whole laboratory was related to the centre plan, rather simply demonstrates the problem of value oscillation. NPEdia and NPDepo, however, present a more subtle picture of how value oscillation is embedded in the more complex layers of multiple infrastructures. To understand this, we must take a closer look at the very concept of both databases as the public library of 'natural product' compounds. As I have shown, this project is officially in line with the wider national science policy programme to establish a public library of chemical compounds, but the uniqueness of this plan lies in its adherence to collecting natural products as opposed to collecting ordinary chemical compounds, as directed in the competing scheme of RIKEN's rival, the University of Tokyo (Fukushima, 2013)[8].

The idea behind this project derives first from the historical fact that natural products have been powerful sources of drug seeds, especially in microbial cases, which have included a variety of powerful discoveries, from penicillin to statins (Newman & Cragg, 2007; Endo, 2006). Second, the relatively strong tradition of Japanese research in this area, to which the antibiotic laboratory belongs, led the promoters to maximize their traditional advantages. Third, the unique chemical structures of these natural products were expected not only to promote the search for drug seeds, but also to lead to the development of unique bio-probes for basic biological research (Fukushima, 2015).

NPEdia was designed to supplement the collection of materials, to serve as a legitimate knowledge infrastructure in the wider context of the drug discovery infrastructure, and to function as an encyclopaedia for natural product compounds with annotated meta-information, such as bioactivity and the details of related assay methods. It was also meant to serve as a catalogue for NPDepo to give details on the further uses of the actual compounds that NPDepo provides

Thus, the NPDepo/NPEdia complex was considered pivotal for the coming chemical biology centre, and a specific team was assigned responsibility for them. This team included a leader – who also acted as office manager of the laboratory administration and coordinator of the other teams that collaborated to develop the various elements of the library/database – as well as a couple of informaticians. This meticulous organization suggests that the managers of this facility require full commitment to its development and administration – demanding high M-B index scores – unlike other team leaders whose attitudes were often lukewarm vis-à-vis such infrastructural obligations.

This initial scheme, however, did not develop as planned, owing perhaps to entangling factors. First, collecting natural products from individual laboratories presented a hindrance because these materials take years or even decades to extract, unlike commercially synthesized materials; being thus the object of researchers' attachment, it is difficult to get them released for public use (interviews, 25/5/2008, 30/6/2011). This aspect can be interpreted as a certain version of value oscillation: researchers officially understand the value of such a library, but they do not want to commit to it. A similar situation was found in the mouse genome database, where young researchers hesitated to submit their research outcomes to the database (Hine, 2006a).

Obtaining materials from retiring researchers before they closed their laboratories was slightly easier, but changing property rights trends, in which universities started to strengthen their control over the products of individual laboratories, are now a problem (interview, 29/5/2014). Thus, a sort of vicious cycle occurred: the failure to collect material enough to demonstrate the merit of such a library with its capacity for processing information in a high-throughput manner further diminished incentives for researchers to submit theirs.

The NPDepo's delay fostered stagnation in NPEdia's development. One of the expected functions of NPEdia, to serve as a full database for natural products, proved too feeble to work autonomously because of competition from rival databases for chemical compounds. Generally, in chemistry, SciFinder, by the American Chemical Society, has been one of the world's most comprehensive and authoritative sources of chemical compounds[9]. The chemists in the laboratory affirmed that SciFinder is sufficient for all parts of their work (interview, 14/8/2014). However, PubChem, released in 2004 by the National Center for Biotechnology Information (NCBI) as an open source database, focuses on the biological activities of small molecules and has recently gained popularity[10]. The informatician in charge of NPEdia explained that in NPEdia's earlier days, the idea of an open database specific to natural product compounds worthy of trial, such as PubChem, was still underdeveloped. However, the speed of data enrichment at PubChem surpassed that of NPEdia, making it extremely difficult for NPEdia to compete with its global rival (interview, 29/5/2014).

However, NPEdia could have retained its advantage if its catalogue function had been developed further. Natural products often occur in minuscule quantities in laboratory settings and are usually very hard to synthesize, which makes their production challenging for synthetic chemists – in some cases, global competition has developed among leading chemists to synthesize natural products first[11].

This situation differs generally from that of chemical or genomic databases. For instance, the chemical databases mentioned above provide ample data related to methods of synthesis or about the vendors that sell such compounds. In the genomic database case, the recent development of commercial service companies has made it possible to quickly produce the necessary vectors from the genetic sequence information in such databases when a researcher asks for them. In other words, there are large networks of articles, laboratories, and vendors between the data in the database and the corresponding materials, which constitute a sub-layer of infrastructure, enabling the users of such databases to adapt the information to develop the materials they need (interviews, 22/5/2014; 6/6/2014; 22/8/2014)[12].

In the case of natural products, this sub-layer has not developed fully, because of limited quantities and difficulty in reproduction. Thus, even if data about a particular compound are gained through the database, the only way to obtain the compound is to ask the laboratory to share the substance. According to a veteran natural product chemist, this is a complicated process because the laboratory may not exist any longer or because the laboratory has such a limited amount of the target compound that it cannot be shared. Even if, in rare cases, the compound might be synthesized and sold by vendors, its purity may be questionable, and further effort to refine the compound by reanalysing its real components may be required (interview, 22/8/2014).

NPDepo would thus be tremendously beneficial for users of such compounds because it would increase the ease of finding the target compound in the library, and the open protocol would simplify the procedure for obtaining material, eliminating negotiations with individual laboratories. NPEdia's full potential would be realized in this way as users could refer to the annotated information within the database and use it as a catalogue, as well.

However, this potential has not been realized thus far. Without NPDepo, NPEdia cannot compete with the existing databases, because its merit is sufficiently strong only with the support of NPDepo. Thus, the apparent powerlessness of NPEdia as a small, emerging database should not be understood solely as the problem of 'gateways' in terms of connecting isolated systems to larger ones (Edwards et al., 2007; Edwards et al., 2009), but also in the context of the data-material complex, where the material scarcity of natural products may have produced a unique set of data-material relations not seen in the wider genres of chemical or genetic databases.

This situation also relates to the relative invisibility of the problem of value oscillation here, in contrast with the preceding case of the chemical biology centre. Needless to say, as part of the centre, NPDepo/NPEdia would inevitably invite value oscillation for those who were obliged to commit themselves to infrastructural work. However, the more visible aspect of value oscillation lies at the sub-layer of the infrastructure – in the supporting network that enables the process of converting data in the database to its corresponding materials, under the guise of individual laboratories' reluctance to submit their materials to support the coming library as infrastructure. This contrasts with the case of the chemical biology centre, where the sway of the M-B index is much more easily observable as the centre scheme itself has more manifestly advanced.

The Virtual Library as the Coming Knowledge Infrastructure?

To observe the knowledge aspect of value oscillation more clearly in the emerging knowledge infrastructure of NPEdia—which has been realized only to a limited degree—let us briefly examine a supplementary case: the emerging virtual library of chemical compounds within the related scheme of RIKEN's drug discovery pipeline. This idea has been promoted by a research group related to the so-called K supercomputer in Kobe, West Japan, as part of a scheme to redevelop the city within a large biomedical complex after the 1995 earthquake (KBIC, 2012). K, from *kei*, meaning 'quadrillion' in Japanese, symbolizes the computation of 10 petaflops per second; this computer is intended to have the fastest computing speed in the world[13]. A number of projects related to this supercomputer are specifically concerned with computational drug discovery. There are at least two major plans: The first is to build a huge library of virtual chemical compounds, and the other to analyse ligand protein interaction using big data[14].

The first plan relies on use of Archem – existing software originally designed for rapid analysis of the optimal paths for synthesizing the target compound – so as to produce large amounts of virtual compounds by reversing the process. The research group succeeded in producing five billion virtual chemical compounds (Ashida, 2010), an astronomical number compared to NPDepo's tens of thousands of compounds or to those of the drug companies, perhaps 10 million at best (interview, 12/5/2012). However, this does not include some of the complex cases of natural products that may often exhibit 3D structural complexity, such as chirality (interview, 2/9/2014). The purpose of this library is to examine the possible interaction between virtual compounds and the target proteins to predict the best-fit cases. The rising expectation that the supercomputer would handle huge computational loads made computer companies like Fujitsu eager to participate in this newly emerging field[15].

However, these methods are not without problems. First, the issue of computational explosion remains in terms of how to balance between calculations based upon either Newtonian or quantum dynamics, and how (not) to calculate the influence of the molecules in the mediating substances, such as solutions or intracellular environments, existing between proteins and their ligands (interview, 12/8/2014). Most problematic, however, is the huge amount of noise. Just as in the past case of combinatorial chemistry, where the once-popular high-throughput production of new compounds has lost its glamour because of the huge nonsensical structures it produces (Barry, 2005; Borman, 2004), the virtual library must also sort significant structures from the huge amount of meaningless ones (interview, 2/9/2014). In fact, past reports indicate that existing calculations not done by K computers have produced a prediction success rate of less than 10% for proper protein-ligand binding (Kanai, 2012).

Thus, the second programme is designed to raise this success rate by enabling the computer to learn the pattern of such bindings using the existing databases on protein–ligand relations. In principle, this is performed similarly to the way a neural computer learns fingerprints or human faces. The success rate for prediction is expected to double from the traditional way of computing the molecular dynamics of these interactions (Kanai, 2012; Okuno, 2012).

In terms of value oscillation, these new radical features reveal intriguing problems not clearly seen in the NPEdia case. Although these programmes are still largely in development and are not ready for public use, their main researchers have enumerated hindrances to plan development, some of which I interpret as indicating value oscillation. For instance, they are uncertain whether they should continue maintenance work to promote the public use of this library as a resource centre after the present phase of system development. The laboratory head responsible for this scheme seems to have high M-B index scores, as he fully recognizes the importance of the infrastructural aspect of his role. His somewhat subtle value oscillation, then, may derive from his peculiar background, a hybrid of computing sciences and other disciplines, such as biology and astronomy. This background often makes him unsure of his position in each of these communities. At informational science meetings, he often finds his colleagues too excited over trivial software innovations; in our terms, he regards them having insufficient M-B index scores for what they should do. However, in life science meetings, he feels that his computational approach is often merely subcontract work for the mainstream wet approaches, meaning that, paradoxically, he is dissatisfied with the way the biology community regards his work as infrastructural.

This means that his generally high M-B index is not sustainable all the time; occasionally, he feels that the essential work that he assigns to his staff – related to deleting all possible noise or the nonsensical chemical structures produced in the library – can be problematic when considering career development possibilities after such tedium (cf. Hine, 2008 for similar uncertainty) (interview, 2/8/2014).

The researcher in the second programme for the machine learning of ligand–protein binding, who seemed to have a very low M-B index score, shrewdly evaded the service aspect of work by entrusting it to a venture company that he established, exempting his laboratory from any further infrastructural work so that he could concentrate on the development of the new method (interview, 8/8/2014). Nevertheless, the value problem is unavoidable when his new method is applied in the real drug discovery context. The problem is how to gain support from chemists for synthesizing their computational predictions into the embodied compounds. The researcher admits that to synthesize the outcome of his very pragmatic machine learning with a process that is theoretically blackboxed would be considered by the synthetic chemists' community to be service work without scientific value – the M-B index score is close to zero here; thus, he asks for help with synthesis only from an old friend from high school. In the case of the virtual library, the researcher from the beginning plans to entrust the job to companies to avoid possible conflicts with scientists who do not want such subcontract-like duties (interviews, 8/8/2014; 2/9/2014).

These programmes are in a development stage wherein their innovative characteristics are spotlighted in public, but eventually, they will move into a maintenance – that is, infrastructural – phase. The knowledge aspects of such infrastructural efforts entail the problem of data-material relations similar to the case of natural products above – namely, the problem of collaborating with synthetic chemists whose M-B index scores are often close to zero in terms of doing service work purely for such a virtual method. Likewise, in natural products, the chemists tend to be hesitant to become involved with the library plan, which is also interpretable as showing a low M-B index score.

This rather unstable relation between the knowledge and related material aspects in the form of non-collaboration by synthetic chemists shows the inherent instability of the in-process knowledge infrastructure. On this point, another specialist in the simulation of protein structures who participates in the K computer programme pointed out the inevitable duality of the infrastructural and innovative aspects of computer technology and the difficulty of balancing them. He noted that computer technology is now widely distributed even in the basic tools of the structural analysis of proteins – namely, in X-rays and NMR spectroscopy – where complex signals are analysed with the help of computerized data processing. Thus, he underscored that even in his laboratory, the aspects of the cutting edge of innovation and infrastructural work exist side by side. He emphasized that the latter should be treated carefully in such ways as developing and fine-tuning the software needed for such simulations, which is ordinarily seen as infrastructural work that does not produce appreciable credentials; thus, careful persuasion is needed to enlist members of the laboratory for such work (interview, 12/8/2014).

Discussion

The main claim of this research is that value oscillation is intrinsic to consolidating and maintaining infrastructures of any type. The problem, then, is how it takes shape and is dealt with according to different degrees of infrastructure consolidation within multiple layers. In fact, a series of preceding discussions have taken place on the perpetual tension or contradiction between the approaches of biology and computer science, often dubbed 'wet' and 'dry' approaches in the research on genomic sciences. For instance, García-Sancho (2012) traces the relation between genetic/protein science and sequence technique as a constant swinging between antagonism and accommodation. In more detailed microscopic studies, Lewis and Bartlett (2013) discuss the problem of bio-informatician identity, which sways between scientists pursuing new knowledge and technicians supporting the jobs of wet biologists. In preceding studies more on knowledge-infrastructural aspects of laboratory work, Star and Ruhleder (1996: 126) refer to a primordial case of value oscillation that they call 'tool building and the reward structure' in their case study of the gene-sequencing network. In relation to the mouse-based genetic database, Hine (2006a) details various potential conflicts and their avoidance in cases similar to the institutional separation between biologists and the resource centre, while her analysis of systematics (Hine, 2008) delineates the value oscillative aspects in more detail.

Compared to these preceding examples, the two cases in this paper appear to be situated on a more complex institutional ladder – one in the chemical biology centre, the other in the K computer project, and both within RIKEN's wider programme. These two are also situated in the wider context of a more established knowledge infrastructure: the databases of chemical and genomic information.

Hence, the phenomenon of value oscillation is most visible at the rather established level of the centre, whose purpose from the beginning has been torn between the goals of an infrastructural service centre and those of a centre for innovating research; the M-B index appeared to be literally fluctuating, as observed in both the leader's and researchers' discourses and action in various contexts. In contrast, the subsequent cases of NPEdia and the virtual library demonstrate a more complex picture owing to their being embedded in multiple layers of both organizations and other databases while falling short of establishing a proper level of autonomy. Hence in the case of NPEdia, the value oscillation is spotted in the sub-layer that supports this database, while in the case of the virtual library, diverse strategies were observed for avoiding a double-bind situation: that is, the shrewd avoidance of a further commitment to maintenance and dissatisfaction with the indeterminate character of the concerned researcher's role in terms of infrastructure development. Thus, compared to the preceding arguments that emphasized a rather black-and-white image of contradictory values, these cases exhibit a more subtle and layered embodiment of value oscillation, as well as diverse ways the practitioners deal with it.

Conclusion

I have argued here that current studies of infrastructure have suffered from historical amnesia lacking critical dialogue with the preceding Marxist discussions on Infrastructure in terms of the genealogy of concepts. I have pointed out that an opportunity has been missed for theoretical dialogue in relation to at least three major questions, the last of which is highlighted in this paper: the pivotal importance of the Marxist operation of inverting the unseen value of Infrastructure, which has been occulted by the shadow of Superstructure – the operation represented by the phrase 'turning Hegel on his head'. I claimed that this operation is intrinsically isomorphic with Bowker's 'infrastructural inversion', now regarded as pivotal in contemporary research on this topic. Behind the need for this operation lies the recognition that the concept of infrastructure is an oxymoron, imbued with contradictory meaning – that is, infrastructure is endowed with power while it is simultaneously inferior to the surface value.

'Value oscillation' is the term used to describe this double-bind situation whereby practitioners hang in indeterminacy between opposite vectors, and the M-B index is the tool used to visualize these oscillating values. Two cases of such value oscillation were taken from the drug discovery infrastructure building in Japan where the knowledge of drugs is uniquely entangled with the physical material in a complex, layered manner.

To elucidate further implications of my approach here, we must note that the recent rise in concern with the infrastructure in the STS community derives not only from growing academic interest in such individual cases as the computer network or energy infrastructure but also from increasing attention within the research community to the more structuralized, *longue duree* elements of socio-technical development rather than to the early and rapidly changing aspects of technoscientic transformation. These renewed intellectual concerns can be observed in such diverse expressions as a reference to the 'cold' situation (Rip, 2010), the 'obduracy' of urban technology (Hommels, 2005) or even 'the shock of the old' (with regard to technologies) (Edgerton, 2006; cf. Fukushima, 2015).

Two points can be drawn from this observation. The first is the merit of talking about infrastructure vis-à-vis the related concepts cited above. Conceptually, infrastructure leads us to focus on the dual aspects of a) its power to exert influence upon that which hinges upon it and b) its invisibility. In this article, I have pointed out the diverse strategies produced by value oscillation, ranging from devotion to shadow work – that is, efforts to raise the status of what is invisible – to the minimal commitment devoted to maintaining and repairing infrastructure. Beyond the micro-sociological examples presented in this paper, larger-scale and more historical consequences of such value oscillation will be similarly important in further examining the longitudinal dynamics of the infrastructure at large – represented, say, by such instances as the recent issue of roads and bridges in decay, owing to politicians' general lack of interest in their proper maintenance (Nemoto, 2011).

The second point is the yet unexamined relationship, in this era wherein STS scholarship highlights the rapidly changing, unstable network of humans and nonhumans, between current concepts of infrastructure and the century-old use of 'structure' in the social sciences. Though not detailed above, the concept of infra-'structure' is not only comparable to its Marxist counterpart but it also partakes of the larger genealogy that has taken its intellectual inspiration from the concept of structure found in structural functional sociology or even structuralism.

Take, for instance, my formulation of value oscillation vis-à-vis the dual aspects of infrastructure – namely, its power to exert influence and its invisibility. If we slightly modify this idea to consider the contrast between the various forms of our 'existence' in terms of l'*engagement*, and the power of invisible structure – whatever that means – to exert an implicit influence upon this existence, this contrast dimly echoes a well-known historical controversy: Sartre's revised concept of existentialism, somewhat modified by his conversion to historical materialism (Sartre, 1976), versus Levi-Strauss's (1966) fatal criticism in which he highlighted the determining power of classification and taxonomy as the invisible structure that regulates our very understanding of history. In fact, the missing link with Marxism which I have highlighted in this paper is only the tip of the iceberg in terms of possible linkages. Godelier (2011; Godelier et al. 1978), for instance, can also be regarded as a concrete embodiment of the confluence of both Marxism and structuralism, which later leads to the thesis of 'classification as infrastructure' that is foundational for the current discussion of knowledge infrastructure.

My emphasis on the resurgence of the intellectual concern with the concept of structure lurking in infrastructural studies, however, does not deny the novelty of the later approach vis-a-vis the earlier. Compared to the more traditional ways of dealing with structure – either as an invisible mental structure or as the social structure, mostly as it relates to humans – current infrastructure studies provide a series of fresh perspectives on the socio-technical complex. Its workings can be more closely observed through modern than through more traditional ways, such as with case studies on roads and databases, which were unanticipated in the past. This is why infrastructure studies, even if they share a concern with past problematics in the social sciences, do not merely repeat the past (cf. Marx, 1994) but may be considered as more finely-tuned re-examinations of persisting controversies from the past, generated by the historical genealogy of concepts through which we see the world.

Notes

1 The earliest version of this formula appeared in the posthumous publication of a draft called *Grundrisse der Kritik der politischen Ökonomi* (Outline of the Critique of Political Economy), written around 1857–58.

2 The labour process theory is a possible candidate for bridging these two research traditions (Braverman, 1974; Nakaoka, 1971; Knights & Willmott, 1990; Sturdy et al., 1992). Vann and Bowker (2006), somewhat exceptionally, reflect this line of concern by focusing on the production side of e-science, emphasising the role of funding agencies, thus recalling classical arguments on the role of the capitalist class.

3 This nuance can be contrasted with technological *regimes*, highlighting the total visibility of the phenomena (Rip, 1995; Rip & Kemp, 1998, followed by many).

4 Included in this category are requests for more attention to neglected aspects of laboratory technicians and technical workers at large (Barley & Bechki, 1994; Barley & Orr, 1997) as well as the Burri's (2008) analysis of radiologists' strategy in terms of cultural capital and boundary work.

5 This paper draws on data from various phases of my ethnographic research in the antibiotic laboratory and chemical biology (2007–2010), the Protein 3000 Project (2010–2013), and drug discovery infrastructure (2013–), which are all related to RIKEN. Interviews were conducted with researchers in various genres on the topic of this theme, both inside and outside the institute.

6 RIKEN's recent programme can be seen at http://www.riken.jp/dmp/english/index_en.html (accessed 22/8/2014)

7 In 2007, ten teams covered the following themes:
1. Streptomyces
2. Genetic analysis of secondary metabolites
3. Fractions
4. Chemical library
5. Compound array
6. Protein analysis
7. Cancer related issues
8. Cell cycles
9. Transcription
10. Chemical compounds at large 2–5 and 10 are largely for infrastructural works

8 See Parry (2004) for the history of the U.S. Natural Products Repository of the National Cancer Institute. Compared to this global scheme, NPDepo is straightforwardly intended for public use for both research and drug discovery.

9 SciFinder's official home page is http://www.cas.org/products/scifinder (accessed 22/8/2014). For its history since 1995, see Chemical Abstract Service (2007).

10 PubChem, launched in 2004 by the National Center for Biotechnology Information (NCBI), is a free-access database focusing on the biological activities of small molecules. It has experienced serious friction with SciFinder (Marris, 2005). https://pubchem.ncbi.nlm.nih.gov/about.html (accessed 22/8/2014).

11 The global competition of synthesizing taxol, an anti-cancer material extracted from Pacific yew, is such a case. In 1993, R. Holton succeeded in its total synthesis. Despite millions of dollars spent, the resulting method, which has more than 40 steps, has not been used for actual drug production (Sato, 2007: 78-86).

12 The case presented here highlights the limit of Parry's (2004) claim concerning the growing use of what she calls the dominance of *ex-situ* data mining. In addition, this line partially refutes Elvebakk's (2006) claim that chemistry has largely become a matter of examining information.

13 http://www.aics.riken.jp/en/k-computer/about/ (accessed 20/8/ 2014).

14 http://www.mext.go.jp/b_menu/houdou/24/09/__icsFiles/afieldfile/2012/09/04/1325265_1_1.pdf (accessed 20/8/2014).

15 Thus, the University of Tokyo, Fujitsu, and the Kowa Company announced the discovery of an anti-cancer drug candidate through computer-based virtual design. http://www.rcast.u-tokyo.ac.jp/research/report/2014/140807PR.pdf (accessed 18/8/2014).

Acknowledgements

I thank all the interviewees in RIKEN and elsewhere for kindly answering my questions. I thank Geof Bowker for introducing me to the world of knowledge infrastructure studies; Antti Silvast and two anonymous reviewers for their constructing comments. This research was conducted with the support of grants-in-aid for scientific research from the Japan Society for the Promotion of Science, under Grant 26350359 (2014–2016).

References

Althusser L & Balibar É (1970) *Reading Capital*. London: New Left Review.

Anderson P (1976) *Considerations on Western Marxism*. London: New Left Books.

Ashida J (2010) Sharing Information Using ARChem. *CICSJ Bulletin* 28(1): 15-16. (in Japanese)

Barley S & Bechky B (1994) In the Backrooms of Science: The Work of Technicians in Science Labs. *Work and Occupations* 21(1): 85-126.

Barley S & Orr J (eds) (1997) *Between Craft and Science: Technical Work in U.S. Settings*. Ithaca, NY: ILR Press.

Barry A (2005) Pharmaceutical Matters: The Invention of Informed Materials. *Theory, Culture & Society* 22(1):51-69.

Bartfai T & Lees G (2006) *Drug Discovery: From Bedside to Wall Street*. Amsterdam: Elsevier Academic Press.

Bataille G (1988) *The Accursed Share: An Essay on General Economy*. Cambridge, Mass.: Zone Books.

Bateson G (1972) *Steps to an Ecology of Mind: Collected Essays in Anthropology, Psychiatry, Evolution, and Epistemology*. Chicago: University of Chicago Press.

Baudrillard J (1981) *For a Critique of the Political Economy of the Sign*. Candor, NY: Telos Press.

Bensaude-Vincent B & Stengers I (1996) *A History of Chemistry*. Cambridge Mass.: Harvard University Press.

Borman S (2004) Rescuing Combichem: Diversity-oriented Synthesis Aims to Pick Up Where Traditional Combinatorial Chemistry Left Off. *C&EN: Science & Technology* 82(40): 32-40

Bowker G (1994) Information Mythology. The World of/as Information. In: Bud-Frierman L (ed) *Information Acumen: The Understanding and Use of Knowledge in Modern Business*. London: Routledge, 231-247.

Bowker G & Star S (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, Mass.: MIT Press.

Bowker G, Baker K, Millerand F & Ribes D (2010) Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: J Hunsinger, J Allen & L Klastrup (eds) International Handbook of Internet Research. Drotrecht: Springer, 97-117.

Braverman H (1974) *Labor and Monopoly Capital: The Degradation of Work in the Twentieth Century*. New York: Monthly Review Press.

Burri R (2008) Doing Distinctions: Boundary Work and Symbolic Capital in Radiology. *Social Studies of Science* 38(1): 35-62.

Carse A (2012) Nature as Infrastructure: Making and Managing the Panama Canal Watershed. *Social Studies of Science* 42(2): 539-563.

Chemical Abstract Service (2007) A National Historic Chemical Landmark. Available at: http://www.acs.org/content/dam/acsorg/education/whatischemistry/landmarks/cas/chemical-abstracts-service-commemorative-booklet.pdf (accessed 28.8.2014).

Clarke A & Fujimura J (eds) (1992) *The Right Tools for the Job: At Work in Twentieth-Century Life Sciences*. Princeton: Princeton University Press.

Edgerton D (2006) *The Shock of the Old: Technology and Global History Since 1900.* Oxford: Oxford University Press

Edwards P, Jackson S, Bowker G & Knobel C (2007) Understanding Infrastructure: Dynamics Tensions and Design. NSF Office of Cyberinfrastructure. Available at: http://cohesion.rice.edu/Conferences/Hewlett/emplibrary/UI_Final_Report.pdf (accessed 10.6.2014).

Edwards P, Bowker G, Jackson S &Williams R (2009) Introduction: An Agenda for Infrastructure Studies. *Journal of the Association for Information Systems* 10(5): 364-374.

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.

Edwards P, Jackson S, Chalmers M, Bowker G, Borgman C, Ribes D, Burton M & Calvert S (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Available at: http://knowledgeinfrastructures.org/ (accessed 20. 4.2014).

Elvebakk B (2006) Network of Objects: Practical Preconditions for Electric Communication. In: Hine C (ed) *New Infrastructures for Knowledge Production: Understanding E-Science*. Hershey: Information Science Pub.,120-142.

Endo A (2006) *The Discovery of a New Drug Statin: Challenging Cholesterol.* Tokyo: Iwanami-shoten. (in Japanese)

Epstein R (2006) *Overdose: How Excessive Government Regulation Stifles Pharmaceutical Innovation.* New Haven: Yale University Press.

Epstein S (2007) *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press.

Fujimura J (1996) *Crafting Science: A Sociohistory of the Quest for the Genetics of Cancer*. Cambridge Mass.: Harvard University Press.

Fukushima M (2005) On Small Devices of Thought: Concepts, Etymology and the Problem of Translation. In: Latour B & Weibel P (eds) *Making Things Public: Atmosphere of Democracy*. Cambridge Mass.: The MIT Press, 58-63.

Fukushima M (2013) Between the Laboratory and the Policy Process: Research Scientific Community and Administration in Japans Chemical Biology. *East Asian Science Technology and Society* 7: 7-33.

Fukushima M (2015) Resilience in Scientific Research: Understanding How Natural Product Research Rebounded in an Adverse Situation. *Science as Culture*. (First on line).

Fukushima M (2016) Constructing Failure in Big Biology: The Socio-technical Anatomy of Japan's Protein 3000 Project. *Social Studies of Science* 46(1): 7-33.

García-Sancho M (2012) *Biology Computing and the History of Molecular Sequencing: From Proteins to DNA 1945-2000*. Basingstoke: Palgrave Macmillan.

Gaudillière JP & Löwy I (eds) (1998) *The Invisible Industrialist: Manufactures and the Production of Scientific Knowledge*. London: Palgrave Macmillan.

Geertz C (1973) *The Interpretation of Cultures*. New York: Basic Books.

Godelier M, Bloch M, Claessen H, Gilmore D, Pi-Sunyer O & Tagányi Z (1978) Infrastructures, Societies and History [and Comments]. *Current Anthropology* 19(4): 763-771.

Godelier M (2011) *The Mental and the Material: Thought Economy and Society.* London: Verso.

Hayek F (1978) *New Studies in Philosophy, Politics, Economics and the History of Ideas*. London: Routledge & Kegan Paul.

Heidegger M (1968) *What is a Thing?* Chicago: H. Regnery.

Hine C (2006a) Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. *Social Studies of Science* 36 (2): 269-298.

Hine C (ed) (2006b) *New Infrastructures for Knowledge Production: Understanding E-Science*. Hershey: Information Science Pub.

Hine C (2008) *Systematics as Cyberscience: Computers Change and Continuity in Science*. Cambridge Mass.: the MIT Press.

Hommels A (2005) *Unbuilding Cities: Obduracy in Urban Sociotechnical Change*. Cambridge Mass.: the MIT Press.

Howard D & Klare K (eds) (1972) *The Unknown Dimension: European Marxism since Lenin*. London: Basic Books.

Hughes T (1983) *Networks of Power: Electrification in Western Society 1880-1930*. Baltimore: Johns Hopkins University Press.

Joerges B & Shinn T (2001) A Fresh Look at Instrumentation: An Introduction. In: Joerges B & Shinn T (eds) *Instrumentation between Science, State and Industry*. Dortrecht: Kluwer Academic Publishers, 1-13.

Kanai C (2012) In Silico Screening Using Chemical Genomic Information. *CICSJ Bulletin* 30(4): 69-72. (in Japanese)

Karatani K (2014) *The Structure of World History: From Modes of Production to Modes of Exchange*. Durham North, Carolina: Duke University Press.

KBIC (2012) Kobe Biomedical Innovation Cluster. Available at: http://www.city.kobe.lg.jp/information/project/iryo/img/iryou.pdf (accessed 20.8.2014).

Keating P & Cambrosio A (2003) *Biomedical Platforms: Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*. Cambridge Mass.: The MIT Press.

Keating P & Cambrosio A (2012) *Cancer on Trial: Oncology as a New Style of Practice.* Chicago: The University of Chicago Press.

Knights D &Willmott H (eds) (1990) *Labour Process Theory.* London: Macmillan.

Kubinyi H (2003) Drug Research: Myths Hype and Reality. *Nature Reviews Drug Discovery* 2: 665-668.

Latour B (2005) From Realpolitik to Dingpolitik or How to Make Things Public. In: Latour B & Weibel P (eds) *Making Things Public: Atmosphere of Democracy.* Cambridge Mass.: the MIT Press, 14-41.

Lévi-Strauss C (1966) *The Savage Mind*. Chicago: Chicago University Press.

Lewis J & Bartlett A (2013) Inscribing a Discipline: Tensions in the Field of Bioinformatics. *New Genetics and Society* 32(3): 243-263.

Lichtheim G (1971) *From Marx to Hegel*. New York: Herder and Herder.

Luhmann N (1980) *Gesellschaftsstruktur und Semantik : Studien zur Wissenssoziologie der modernen Gesellschaft*. Bd.1. Frankfurt a. M.: Suhrkamp.

Marris E (2005) Chemical Reaction. *Nature* 437(October): 807-809.

Marx K (1973) *Grundrisse: Foundations of the Critique of Political Economy (rough draft)*. Harmondsworth: Penguin Books.

Marx K (1976-1981) *Capital: A Critique of Political Economy*. London: Penguin Books in association with New Left Review.

Marx K (1994) *The Eighteenth Brumaire of Louis Bonaparte*. New York: International Publishers.

Mazzucato M (2013) *The Entrepreneurial State: Debunking Public vs. Private Sector Myths.* New York: Anthem Press.

McLellan D (1979) *Marxism after Marx: An Introduction*. London: Macmillan.

Mody C (2011) *Instrumental Community: Probe Microscopy and the Path to Nanotechnology*. Cambridge Mass.: the MIT Press.

Nakaoka T (1971) *The Philosophy of Factory*. Tokyo: Heibonsha. (in Japanese).

Nemoto Y (2011) *Decaying Infrastructure: Another Crisis that Creeps Up to Us*. Tokyo: Nihonkeizai Shinbunsha. (in Japanese)

Newman D & Cragg G (2007) Natural Products as Sources of New Drugs over the Last 25 Years. *Journal of Natural Products* 70(3): 461–477.

Okuno K (2012) Chemo-genomics: Form Genome to Chemical Space. *CICSJ Bulletin* 30(4): 60. (in Japanese).

Parry B (2004) *Trading the Genome: Investigating the Commodification of Bio-Information*. Princeton: Princeton University Press.

Petryna A (2009) *When Experiments Travel: Clinical Trials and the Global Search for Human Subjects*. Princeton: Princeton University Press.

Polanyi K (1944) *The Great Transformation*. New York: Farrar & Rinehart.

RIKEN (2005). *88 Years of RIKEN's History* (in Japanese). Available at: http://www.riken.go.jp/r-world/info/release/riken88/book/ (accessed 22.1.2007).

Rip A (1995) Introduction of New Technology: Making Use of Recent Insights from Sociology and Economics of Technology. *Technology Analysis and Strategic Management* 7: 417-31.

Rip A (2010) Processes of Entanglement. In: Akrich M, Barthe Y, Muniesa F & Mustar P (eds) *Débordements: Mélanges offerts à Michel Callon.* Paris: Presses des Mines, 381-392.

Rip A & Kemp R (1998) Technological Change. In: Rayner S & Malone E (eds) *Human Choice and Climate Change*. Columbus, Ohio: Battelle Press, 327-399.

Rheinberger HJ (1997) *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford, CA: Stanford University Press.

Rydzewski R (2008) Real World Drug Discovery: A Chemists Guide to Biotech and Pharmaceutical Research. Amsterdam: Elsevier.

Sahlins M (1972) *Stone Age Economics*. Chicago: Aldine.

Sartre JP (1976) *Critique of Dialectical Reason*. London: New Left Books.

Sato K (2007) *Welcome to a Museum of Organic Chemistry.* Tokyo: Gijutsu-hyoronsha. (in Japanese).

Shiozawa Y (2002) *The Legacy of Marx: From Althusser to Complexity*. Tokyo: Fujiwara-Shoten. (in Japanese)

Star S (1999) The Ethnography of Infrastructure. *American Behavioral Scientist* 43(3): 377-391.

Star S & Bowker G (2002) How to Infrastructure? In: Lievrouw L & Livingstone S (eds) *The Handbook of New Media: Social Shaping and Consequences of ICTs*. London: Sage Publications, 151-162.

Star S & Ruhleder K (1996) Steps toward an Ecology of Infrastructure: Borderlands of Design and Access for Large Information Spaces. *Information Systems Research* 7(1): 111-134.

Sturdy A, Knights D & Willmott H (eds) (1992) *Skill and Consent: Contemporary Studies in the Labour Process*. London: Routledge.

Ueno A (2008) Study on the Organizational Knowledge of Research Laboratories with Long Histories. Discussion Paper No.50, National Institute of Science and Technology Policy (NISTEP): 1-94. (in Japanese)

Vann K & Bowker G (2006) Interest in Production: On the Configuration of Technology-Bearing Labors for Epistemic IT. In: Hine C (ed) (2006) *New Infrastructures for Knowledge Production: Understanding E-Science*. Hershey: Information Science Pub., 71-97.

Vertesi J (2014) Seamful Spaces: Heterogeneous Infrastructures in Interaction. *Science Technology & Human Values* 39(2): 264-284.

Weick K & Westley F (1999) Organizational Learning: Affirming an Oxymoron. In: Clegg S, Hardy C & Nord W (eds) *Managing Organizations: Current Issues*. London: Sage Publications, 440-458.

Wikstrom M (2007) Chemical Biology in the USA. Swedish Institute for Growth Policy Studies. Available at: http://www.vinnova.se/upload/dokument/Verksamhet/Bioteknik/Rapporter_Life_Science/Chemical%20 Biology.pdf (accessed 21.1.2012).

Williams R (1976) *Keywords: A Vocabulary of Culture and Society.* London: Fontana/Croom Helm.

Building Knowledge Infrastructures for Empowerment: A Study of Grassroots Water Monitoring Networks in the Marcellus Shale

Kirk Jalbert

FracTracker Alliance / jalbert@fractracker.org

Center for Science, Technology and Society, Drexel University, USA

Abstract

This paper characterizes the activities of two nongovernmental environmental monitoring networks working to protect watersheds in the Northeast United States from the impacts of shale oil and gas extraction. The first is a grassroots coalition of advocacy groups. The second is a large network managed by academic institutions. In both cases, knowledge infrastructures were built to distribute resources and to assist members in using data to make scientific claims. I find that the designs of these knowledge infrastructures can reproduce entrenched dynamics of power in ways that advance the agendas of some stakeholders more than others. However, findings also suggest that the 'grassroots' of infrastructures can tactically alter power relationships and redistribute resources to their advantage. By bringing a discourse of power and empowerment into the study of knowledge infrastructures, this paper offers a theoretical contribution to better understand the conditions by which marginalized stakeholders shape knowledge work to deal with complex scientific and environmental problems.

Keywords: knowledge infrastructures, public empowerment, citizen science, environmental justice

Introduction

Across the United States, energy companies are drilling for oil and natural gas using often disputed methods of extraction known as hydraulic fracturing—a drilling technique that injects million of gallons of water and chemical additives into a well to release hydrocarbons from underground shale rock formations. Complicit in this process is the use of horizontal drilling, which allows operators to extend their reach as far as two miles in any direction (US Energy Information Agency, 1993). In addition to tens of thousands of wells that have been "fracked" in the U.S., recoverable shale oil and gas deposits can also be found throughout North and South America, Europe, Africa, Asia, and Australia (US Energy Information Agency, 2013).

A growing body of evidence suggests that, amongst other environmental threats, watersheds in the vicinity of hydraulic fracturing operations can be impacted by numerous extraction-related problems including seepage from damaged gas well casing, improper waste disposal, trucking accidents, and underground migration of drilling fluids (Donlin, 2010; Entrekin et al., 2011;

Llewellyn et al., 2015). Assessing these impacts is complicated by the fact that hydraulic fracturing is a poorly regulated practice. In the U.S., drilling companies are largely exempt from the Safe Drinking Water Act and the Clean Water Act (Soraghan, 2011). Some states do not require companies to disclose records of chemicals used in drilling; others lack timely systems to notify the public of regulatory violations (Malone et al., 2015).

Due to this lack of transparency, efforts to understand hydraulic fracturing's impacts have largely fallen on the shoulders of the civil society (nongovernmental) sector—academic researchers, non-profit advocacy organizations, citizen scientists, and concerned citizen groups. In the Marcellus Shale, one of the most actively drilled formations in the Northeast U.S., civil society groups have established surface water monitoring programs to assess potential changes in water quality that might result from oil and gas extraction. Beginning in 2010, a number of capacity building organizations (typically nonprofits that provide services to local environmental groups) developed sampling and quality assurance plans to assist concerned citizens in measuring basic water quality indicators (Jalbert et al., 2014; Kinchy & Perry, 2012). Training programs were organized to propagate standardized protocols and to establish larger monitoring networks. Private foundations provided funds to purchase equipment for these groups, ranging from $100 handheld pocket meters to $1,000 automated "data logger" devices (Jalbert & Kinchy, 2015). Meanwhile academic researches grew interested in aggregating rapidly accumulating data for long-term ecological assessments and watershed-wide geospatial mapping projects.

While civil society water monitoring programs have been active in the region since the early 1970s, this groundswell of new monitoring efforts that emerged in response to shale oil and gas development is significant. In only five years, this field matured from a dispersed collection of projects into a vast community of stakeholders accumulating social capital and technical resources to collect data and ask meaningful questions. The people who invested in these efforts believed that, by generating their own science, they would be empowered to participate in public debates and influence regulatory decision-making.

Knowledge Infrastructures for Civil Society Science

In the science and technology studies (STS) literature, scholars have argued that civil society science groups can alter the balance of power between at-risk communities, regulatory agencies, and polluting industries by developing the means to generate independent knowledge (Corburn, 2005; O'Rourke & Macey, 2003; Ottinger, 2009; Overdevest & Mayer, 2008). This research also illustrates how grassroots monitoring groups can overcome barriers of scientific legitimacy by forming partnerships with experts in professional organizations and academic institutions (Lave, 2012; Morello-Frosch et al., 2005; Savan et al., 2003; Wagenet & Pfeffer, 2007). Such partnerships can coalesce resources, improve data collection methods, open doors to laboratories, and enlist specialists who assist groups in solving technical issues. In this paper, I argue that civil society groups in the Marcellus Shale, when assembling resources and building broader alliances for water monitoring research, also constructed distinct "knowledge infrastructures" (KIs) to question scientific assessments often dominated by powerful institutions.

Foundational research in STS on infrastructures focused on the construction of large-scale development projects, such as electrical grids and transportation projects (Bijker et al., 1987; Hughes, 1987), as well as on computing systems that support cooperative work environments (Bowker et al., 2010; Edwards et al., 2009; Star & Ruhleder, 1996). Recent scholarship has sought to identify and understand the mechanisms of knowledge production, where the "internetworks of people, artifacts, and institutions which generate, share, and maintain specific knowledge about the human and natural worlds" come together (Edwards et al., 2013: 23). While important aspects of water monitoring infrastructures include the wide array of monitoring protocols, data collection tools, and data management systems in use—topics dealt with extensively in prior publications (Gouveia et al., 2004; Pfeffer & Wagenet, 2007; Jalbert et al., 2014; Jalbert & Kinchy, 2015)—the emphasis of this paper is on the social side of knowledge infrastructures; on the relations of people and organizations that define research partnerships.

One of the core concepts in KI research explored in this paper suggests that infrastructures stabilize and become rigid in their maturity. Converging designs can push out other competing standards, instrumentations, and organizational structures, thus fusing how an infrastructure's stakeholders share resources and political power (Pinch & Bijker, 1987; Star & Ruhleder, 1996). Edwards et al. (2013: 13) note that, when this occurs in KIs, it can have "significant distributional consequences, advancing the interests of some and actively damaging the prospects of others." The KI literature has investigated how these struggles occur when defining the meanings of knowledge and data used by stakeholders (Borgman et al., 2012; Bowker & Star, 1999; Edwards et al., 2011). Others have looked at how an infrastructure's intended functions can "break down" and reveal inner tensions (Ribes & Finholt, 2009; Star, 1999).

These are important developments in figuring out the nature of knowledge work, but there remains a gap in understanding relationships of power and empowerment in these struggles. I argue in this paper that KIs can remain amazingly dynamic spaces where power is continually negotiated. An infrastructure's stability can become susceptible to competing demands when the marginalized, peripheral, or what I refer to here as the "grassroots," of infrastructures forward objectives that differ from those who are perceived to have control over infrastructural development. Tactical resistance seeks to change how an infrastructure works while also keeping the core of the infrastructure functionally intact. In the case of research infrastructures built to make sense of environmental pollution, acts of resistance are, at their core, also struggles to build capacity for dealing with real life injustices.

Assessing Power and Empowerment in Knowledge Infrastructures

Bowker et al. (2010: 106) have argued "if participants have been active in the formation of infrastructure elements, they are more likely to have a deeper awareness of alternatives and have had a voice in mediating choices inherent to issues such as standards formation and community goals." The nature of participation and what it means to have voice in infrastructure building is, however, not well understood. STS researchers have developed a robust language to describe new forms of participatory research including citizen science, community-based science, street science, and crowdsourcing science (Corburn, 2005; Fischer, 2000; Irwin, 2001; Moore, 2006). Each of these seeks to illustrate the ways in which professionals and nonprofessionals negotiate power at different stages of research.

One of the more prominent models to emerge comes from the natural sciences and is offered by Shirk et al. (2012). At one end of their spectrum of participation are "contributory" projects, where volunteers collect data for scientists but otherwise have little control over the nature of research. On the other end are "co-created" projects, designed in equal partnerships, that emphasize shared decision-making. Models for evaluating participation are relatively absent from KI studies, but could clarify how KIs engender certain liabilities for less powerful grassroots groups, particularly when they must relinquish control over their work in order to participate in larger research programs.

Similar to poorly defined metrics of participation in infrastructure building, assessments of empowerment—the increased capacity of an infrastructure's stakeholders to design, implement, and evaluate mechanisms that improve their standings in the world—is also weakly defined. Here, I look to the contributions of critical geography, which has a long history of appraising empowerment in knowledge construction projects, particularly in the use of geographic information systems (GIS). For instance, Corbett and Keller (2005a, 2005b) make a distinction between empowerment—"a tangible increase in social influence or political power"—and empowerment capacity—"aspects of the deeper process of change in the internal condition of an individual or community that influence their empowerment" (Corbett & Keller, 2005b: 28). They suggest that catalysts for empowerment can come from gaining access to new information, learning new technical skills, or developing fresh political strategies. Their framework makes a further distinction in that empowerment and empowerment capacity can evolve differently at the scale of the individual versus that of the larger community.

In the remainder of this paper I familiarize assessments of participation and empowerment with KI studies through an examination of two civil society water monitoring networks operating in the Marcellus Shale. The first, built by a coalition of concerned citizen groups called the New York Water Sentinels, was formed in 2011 and later expanded through loose affiliations with the Sierra Club (one of the largest and oldest environmental advocacy organizations in the United States). With an annual budget of only $20,000, roughly 150 volunteers now monitor streams in twelve counties across New York State. The second monitoring network emerged from a project called Three Rivers QUEST (3RQ). 3RQ is supported by $1.3 million in grants awarded to West Virginia University's Water Research Institute by a nonprofit foundation. 3RQ supports a variety of water monitoring programs across the states of Pennsylvania, West Virginia, Maryland, and Ohio.

I find that the KIs built by these two research communities are similar in terms of the resources they offer to their affiliated stakeholders. They differ in the extent to which grassroots members retain control over research agendas at various points of KI development. Relationships of power are found to heavily depend on adopted models of participation that can either aggregate or distribute power, authority, and expertise. I also show that KIs can elude stability and change over time when marginalized groups develop tactics to influence the direction of scientific research. These findings bring to the forefront the importance of evaluating the attributes of participation and empowerment when assessing the long-term affordances of KIs.

Data Sources and Research Methods

Data supporting this paper was gathered from 2011 to 2015 and draws from more than 30 semi-structured interviews conducted with a wide range of stakeholders including representatives of water monitoring networks, government agencies, capacity building organizations, nonprofit watershed groups, academic institutions, data management projects, and major funding foundations. The organizations in this paper are identified accurately, however pseudonyms are used in place of people's real names in order to protect personal privacy. All interviewees were granted the right to review their quotations for clarity and context prior to publishing.

Additional data came from more than 1,000 hours of participant observation with groups mentioned in this paper and many others. A great deal of time was spent on the ground with concerned citizens as they trained to participate in water monitoring projects, collected samples in the streams, and analyzed their data. When studying capacity building organizations, participant observation occurred during visits to their offices, by attending strategy sessions, and by being present at regional summits where outreach coordinators interacted with their constituents.

A significant amount of information was also acquired through embedded or "engaged" research activities that emerged organically when interlocutors asked for assistance with their efforts. As an appointed member of the New York Water Sentinels' science advisory committee, I was able to join weekly planning calls and offer insights from the research. Other discernments came from coordinating quarterly meetings of the "Water Quality Data Coordinator Group" in 2014 and 2015. These gatherings brought together more than twenty representatives from across the Marcellus Shale water monitoring community to build data sharing synergies. Finally, insights offered in this paper came from my experiences as a visiting researcher from 2012 to 2015 with the FracTracker Alliance—a Pennsylvania-based nonprofit that works to enhance public understanding of oil and gas extraction through interactive maps, data analysis, and articles—where I am now the Manager of Community Based Research and Engagement.

Dilemmas of Democratic Governance: The New York Water Sentinels

"You can see it's kind of gurgling," the person next to me commented as we peered over the edge of an access hatch to a nearly 40-foot tall vat of stewing sludge. The smell was overwhelming and we felt a bit uneasy about the rope and emergency flotation device hanging beside us on the railing. This particular tower was but one in a complex arrangement of pipes, pumps and tanks that processed the regular flow of leachate (liquid waste outflows) from the nearby Steuben County Landfill in the Village of Bath, located in upstate New York (Figure 1). A dozen people stood below us on the next platform, listening intently as the plant manager described how drainage from the landfill entered the system on one end, gets piped through the Village of Bath's sewage system, and is eventually discharged into the Cohocton River, a tributary of the upper Susquehanna River watershed.

The Steuben County landfill is the site of a decades-old township dump, originally constructed without a proper leachate treatment system. In 1995, when the landfill sought expansions, the New York Department of Environmental Conservation (NYDEC) assisted Steuben County in building a treatment plant that would not only have the capacity to process leachate from this landfill, but also wastewater from other sources in the state (Hardman, 2014). One growing market for waste processing came from Pennsylvania's Marcellus Shale oil and gas drilling industry.

Despite New York State's 2008 drilling moratorium, and the more permanent ban on hydraulic fracturing enacted in June 2015, a recent report calculated that more than 460,000 tons of solid waste and 23,000 barrels of liquid waste from shale drilling operations in Pennsylvania were processed in New York (Moran, 2015). Facilities accepting this waste included the Chemung County Landfill, Casella Waste Systems, Seneca Meadows Landfill, Allied Waste Systems, Hyland Facility Association, and the Hakes Landfills. Among these facilities, the Chemung County Landfill has accepted the most solid waste, at nearly 200,000 tons. Hyland Landfill and Hakes Landfills have each accepted over 100,000 tons of drill cuttings. These amounts do not include the tens of thousands of tons of drill cuttings used as "daily cover"—a layer of compressed soil placed on top of a landfill at the end of each day.

Drill cutting are highly valued by landfill operators. Their density takes about one-fourth the space of conventional waste, but can be charged at the same per-ton disposal fee. This new revenue stream persuaded operators to expand

Figure 1. A tour the Steuben County leachate pretreatment plant

their facilities in tandem with Pennsylvania's expanding shale oil and gas industry. In one case, Chemung County landfill received approval from the NYDEC to expand their volume of accepted waste from 120,000 tons to 180,000 tons per year, and then began diverting less-valuable county waste to other landfills (Mantius, 2013).

A number of loopholes allow shale oil and gas waste to travel into New York even though much of the gas industry's practices are otherwise limited. According to NYDEC regulations, drilling muds are not considered hazardous waste (NYS Department of Environmental Conservation, 2006). Drilling waste is also exempt from New York's Low Level Radioactive Waste Laws that govern Naturally Occurring Radioactive Materials (NORMs) such as uranium, radium 226, and radium 228. Many residents feel this rule fails to acknowledge that Marcellus Shale drill cuttings are known to contain low levels of radioactive materials (Puko, 2013).

The Steuben County landfill, the site of our tour, did not directly accept drilling waste of any kind from the Marcellus Shale, but one fact was known about the facility—its overbuilt wastewater treatment plant generated revenue for the county by accepting excess leachate from neighboring landfills, including more than 2.2 million gallons worth from Hyland landfill, and nearly 2 million gallons from Hakes Landfill between July 2012 and April 2013 (Mantius, 2013). These facts had many residents in the county worried about the safety of public drinking water supplies and nearby watersheds.

Building Grassroots Infrastructures

Our tour of the Steuben County landfill treatment plant was organized by a local chapter of the New York (NY) Water Sentinels—a grassroots coalition of environmental advocacy groups that began baseline monitoring in watersheds along New York State's border with Pennsylvania in 2011 where Marcellus Shale drilling was expected to occur if the state's moratorium were lifted. The origins of NY Water Sentinels can be traced back to the Concerned Citizens for Cattaraugus County (CCCC), an organization that has worked for years on issues ranging from stopping large windmill farms near homes to opposing new landfills. As part of their initiative to address shale oil and gas issues in New York, the regional Atlantic Chapter of the Sierra Club (representing Northeastern U.S.) became interested in supporting a water monitoring program. The Sierra Club approached the CCCC and provided a seed grant through its National Water Sentinels program to assist members in acquiring equipment. Together they scheduled training sessions with the Alliance for Aquatic Resource Monitoring (ALLARM), an outreach program of Dickinson College in Carlisle, Pennsylvania, that was instrumental in developing volunteer-based shale oil and gas water monitoring protocols across the Marcellus Shale starting in 2010.

Over the next two years the NY Water Sentinels brought on additional volunteers by canvassing at town meetings and local newspapers, as well as by partnering with other regional environmental organizations. Their monitoring network now extends into the watersheds of thirteen counties along the New York and Pennsylvania border where 160 volunteers have made more than 1,500 visits to document conditions at 125 stream sites. I asked Miles Coolidge, an advisor from the Sierra Club Atlantic Chapter who sits on the Steering Committee, about how the NY Water Sentinels evolved. In succinct terms, he described the birth of a grassroots knowledge infrastructure:

> The first year we spent a lot of time getting the QA/QC to work. We built the technical infrastructure. The second year we worked on getting the coordinator groups working—the social infrastructure. Now we need to do more outreach into challenging areas, to develop that sense of community. Our value is to work at the local level. We have to make sure we are embedded in the community.

The NY Water Sentinels have no paid staff or dedicated facilities. Its governing system is one of overlapping committees populated by volunteers. The Steering Committee is the executive body, and is responsible for managing the infrastructure's broader mission. This includes establishing new affiliations with outside partners, determining where future training will occur, making changes to monitoring protocols and how data will be used, as well as deciding what political or legal initiatives they may initiate. Steering Committee members are elected annually by Water Sentinels chapters, but also consist of members of the Sierra Club Atlantic Chapter office who serve an advisory role. Day-to-day governance of the NY Water Sentinels falls upon the Coordinators Committee. Its purpose is to implement the directives of the Steering Committee and, in the process, maximize inclusion of the network's volunteers by soliciting input on monitoring strategies. The Coordinators Committee meets weekly by phone to discuss topics ranging from equipment maintenance, data management issues, quality assurance updates, and the status of their finances.

Other working groups that meet on a semi-regular basis include an External Communications Committee, a Finance Committee, a Fundraising Committee, and a Data Management Committee—all populated by volunteers in the network. The NY Water Sentinels also retain the help of many outside experts who assist the Steering Committee and Coordinator Committee with different tasks. No less than three practicing attorneys advise the Legal Committee. The Science Committee (on which this author sits) regularly consults with professors of biological science, geology, and environmental studies at different universities.

The Tradeoffs of Empowerment

The NY Water Sentinels created what Shirk et al. (2012) might call a "co-created" partnership structure, one that empowered individuals on the front lines to influence the design of their KI. However, these same egalitarian systems also made the network vulnerable to internal friction and competition for resources. The story of how this unfolded within the NY Water Sentinels began at the Hyland Landfill in Wellsville, New York.

In 2013 Casella Waste Systems applied for a permit to expand Hyland Landfill's annual volume of accepted waste by more than 60% in order to accommodate Pennsylvania's drilling cuttings (Donohue, 2015). When this became public knowledge, two members of the local NY Water Sentinels chapter felt it was important to begin monitoring around the landfill as well as its wastewater treatment plant. "We didn't find

Figure 2. Water sampling at a leachate treatment plant outflow pipe

any elevated radioactivity from the Wellsville Water Treatment plant discharge, but we did find elevated radioactivity in a stream running off from the landfill at a designated outfall," Gavin Erwitt, the chapter's coordinator, explained.

As word spread about the possible radioactive risks of drilling waste entering the state, other NY Water Sentinels volunteers began monitoring the outflows of landfills and wastewater treatment plants in their region (Figure 2). Increased attention to landfills quickly became a contentious issue. Some members felt that chasing landfills compromised carefully laid plans for conducting watershed-wide baseline studies. Other concerns were raised as to whether or not taking an advocacy-oriented position would undermine the NY Water Sentinels' ability to raise funds from science-minded benefactors. By contrast, proponents of landfill monitoring were quick to point out that the industry may apply enough pressure to reverse the recent ban on high-volume hydraulic fracturing if natural gas prices rebounded from current record lows. In their view, the ban created an opportunity to extend their mission and take action against other sources of known pollution.

A heated discussion in a Leadership Committee meeting highlighted the character of this debate. "The assumption of our original mission was that fracking would eventually come to New York and we needed to prepare for it," a committee member commented.

A second committee member had a different perspective. "We did originally say that we needed the data to understand baseline. But we need to have the conversation that, now that we have data and we know more, should we expand our mission?" They continued, "Getting people involved at the community level in water quality is an important facet of making a difference. We always incorporated that as part of our discourse. But we haven't been as active in that area as we could be. So the efforts around the Hyland landfill illustrate to me how important outreach is to these areas."

"Well, we should also have a discussion of how this changes the political posture of our groups," a third committee member responded, "Our group has a history of 'opposing' things. Baseline monitoring allowed us to get involved in something that was more objective and positive, to just look at possible violations. Landfill monitoring would bring us back to being a bad-boy watchdog. This is a conversation each group will have to have."

These debates would continue in following months while most NY Water Sentinels chapters continued their usual baseline monitoring work. Nevertheless, volunteers that did take up the cause of landfill monitoring would eventually influence allocations of resources in their infrastructure and draw in outside allies. A dedicated protocol for landfill monitoring was developed with the help of staff at ALLARM, volunteers used funds to purchase equipment for collecting samples around treatment plants, and a nearby laboratory offered to process these samples at cost.

These subtle shifts in resource allocations show some of the benefits of KIs that emerge from partnerships with equal power sharing. When the NY Water Sentinels program began, it emphasized the importance of doing science at the local level. Its resulting governing system respected input from its individual members when steering their KI's development. Some monitoring groups expanded their capacity for empowerment by having the freedom to address new problems discovered in the course of water monitoring. However, for a KI with limited resources, increasing the empowerment capacity of some came at a cost to others with different ideas about how to influence environmental debates. The story of landfill monitoring reveals how KIs can be fluid things and internal power can shift due to the levers of democratic governing systems.

Institutionalizing Grassroots Infrastructures

KIs can also change in moments of vulnerability, such as when resources become scarce, or when citizen scientists struggles for legitimacy. Stakeholders may appeal to powerful institutions for financial, political, or technical support. In these instances, KIs can become susceptible if strengthening alliances with institutions also means giving up control in deciding how KIs function.

By the summer of 2014, the NY Water Sentinels had grown into a formidable presence in New York State. Member groups planned to monitor as many as 150 sites and would offer two new training programs, which they estimated would attract 30 new volunteers. These and other programming expenses, such as public outreach events and laboratory analysis, would be covered by a 2015 budget of roughly $22,000. Unfortunately, the Sierra Club eliminated funding for the National Water Sentinels program at the end of the 2014 fiscal year—a fund that underwrote more than 60% of the NY Water Sentinels' annual budget.

Loss of a major funding source meant that the NY Water Sentinels had to find a way to sustain their hard-won research program by other means. One possibility was to maintain the status quo as a semi-autonomous affiliate of the Sierra Club. They had learned in the past, however, that obtaining funding for local projects while under the umbrella of a large environmental nonprofit could be difficult. Another solution was to become a wholly independent nonprofit, leaving behind their long-time Sierra Club benefactors. This idea did not sit well with those who had been members of the Sierra Club long before the NY Water Sentinels came into existence. Sierra Club advisors suggested a third possibility: The NY Water Sentinels could become an official subprogram of the Sierra Club Atlantic Chapter. The Steering Committee weighed their options and, in late 2014, elected to officially affiliate with the Sierra Club Atlantic Chapter.

Being part of the Sierra Club allowed the NY Water Sentinels to apply for new funding, but the decision had other implications as well. Under their new charter, individual monitoring groups would be required to report their activities in Steering Committee meetings as before, but also now had to report to the Sierra Club Atlantic Chapter's Conservation Chair. Groups could make recommendations to undertake a legal action, but were not permitted to act independently without a detailed review by the Sierra Club. Financially, the NY Water Sentinel's assets would be held in a Sierra Club Foundation account. How these changes might affect the day-to-day operations of the NY Water Sentinels largely remains to be seen. When I asked one chapter coordinator how his group felt about these changes, I was told:

> Look, we have meat eaters and hunters and we have—most of our people are Republicans. They don't have any sympathy for the Sierra Club. They are not members of the Sierra Club. I joined the Sierra Club just so I could do this. I wasn't a member of the Sierra Club. There is a difference between grassroots environmentalism, which is what we do, and aesthetic environmentalism, which is what the Sierra Club does. Our interest is in protecting our backyards. We are NIMBYs [Not In My Back Yard] and we wear the NIMBY badge with honor. The Sierra Club Atlantic Chapter, of people that are involved here, are paid staff. They seem to have little understanding of what it takes to organize and maintain an entirely volunteer group.

These comments reflect how building closer ties to the Sierra Club made some NY Water Sentinel members uncomfortable with the new arrangement. Individual volunteers who had invested time and resources in building the NY Water Sentinels' KI wanted to have a say in its daily operations and broader objectives. Some groups, like the CCCC, also had a long history working in their communities and understood what it took to bring people together and sustain their membership. These groups resented the notion that the Sierra Club might dictate organizing tactics, or narrow how they utilized water monitoring for scientific, political, legal, or other purposes. At present, the Steering Committee and the Leadership Committee are looking for ways to retain greater autonomy while part of the Sierra Club Atlantic Chapter. For instance, they are working to refine the Sierra Club's 501c3 tax reporting (required by federal law to designate non-profit expenditures) so funders can tailor their donations for use by individual monitoring groups to work on specific projects.

Uneasy Alliances in Rigid Infrastructures: Three Rivers Quest

In a different region of the Marcellus Shale, watersheds along southwestern Pennsylvania's border with West Virginia traverse some of the densest coal and natural gas mining fields in the United States. In 2009, researchers from the West Virginia Water Research Institute (WVWRI), based at West Virginia University, began to notice that levels of total dissolved solids (TDS) (a measure of water salinity) in the region's watersheds were exceeding U.S. Environmental Protection Agency (EPA) secondary drinking water standards, particularly in tributaries of the Monongahela River. WVWRI researchers deduced that excess TDS was likely coming from coal and gas extraction sites, but they knew little about where and when pollution discharges were occurring.

The Monongahela River flows from West Virginia into Pennsylvania, and eventually meets the Allegheny River to form the Ohio River in the city of Pittsburgh. Experts in the region agreed that a coordinated strategy was needed to bring together the many individual monitoring programs in these three large watersheds. In 2011, the Pittsburgh-based Colcom Foundation awarded the WVWRI $60,000 to establish what would become known as Mon River QUEST (Quality Useful Environmental Study Teams), a program to aggregate and analyze data from regional watershed protection groups. WVWRI received an additional $750,000 from Colcom in 2012 to expand the program into the Allegheny and Ohio River watersheds—Mon River QUEST was renamed Three Rivers QUEST, or 3RQ. 3RQ received a third grant from Colcom for $500,000 in 2013 to develop the "QUEST Data Management Tool" for storing, managing, and mapping water quality data (West Virginia University, 2013).

The KI built by 3RQ is a complicated arrangement of organizations and institutions with varying resources and objectives. Early on in the project's design, WVWRI identified three research partners to take stewardship over the different watersheds: Wheeling Jesuit University was assigned the Ohio River, Duquesne University to the Lower Monongahela and Lower Allegheny Rivers, and the Iron Furnace Chapter of the Pennsylvania Council of Trout Unlimited (a state-wide network of sport fishing enthusiasts) to the Upper Allegheny River. WVWRI would continue to oversee the Upper Monongahela River. 3RQ's research partners were responsible for collecting bi-weekly water monitoring samples in their respective watersheds. Each research partner also supervised dispersing competitive $3k-$5k mini-grants to independent watershed groups in their territories. Beneficiaries of these grants ranged from small volunteer groups to large watershed associations with dedicated staff. In total, 3RQ would bring together a monitoring network of some 30 groups collecting samples at more than 300 stream sites in four U.S. states.

3RQ's research partners stressed that the diverse research program they had developed would bring resources and expertise to bear on problems important to local watershed groups. "3RQ provides a unique opportunity for academic scientists to engage in community based participatory research—that is, water quality issues identified by our community partners helps to prioritize our research efforts," a researcher at one of the partner universities commented in 2013 (West Virginia University, 2013). Research partners further argued that community groups would be empowered by co-designing research with scientific experts. In practice, however, research partners dictated how 3RQ's KI functioned. This affected everything from resource allocations to how data was used when making scientific claims and political statements. Meanwhile, mini-grantees who came on board with 3RQ expected to have decision-making powers in their participation. Ultimately, disconnects between 3RQ's founding principles and how the project actually functioned would have major implications for the stability of 3RQ's KI.

Erecting Boundaries of Power and Expertise

3RQ's research partners possessed a great deal of power when dealing with local watershed groups. One expression of this power was revealed in the process of determining which groups would receive 3RQ mini-grants, which became important resources for bolstering underfunded and understaffed programs in the region. Lisa Greenfield, a watershed specialist in West Virginia, recalled why her organization applied for a mini-grant in saying, "We were really driven by the need for staff support, and not by the resources that the program was going to offer beyond that." Lisa's group received funding from 3RQ and became part of the monitoring network.

Not all watershed groups were as lucky. Mandy White, a watershed specialist in a Pennsylvania-based organization, recalled having a different experience. Mandy's organization managed a large network of automated data loggers funded by Colcom. She assumed her organization would be supported by 3RQ:

> Colcom let us know that, because of this relationship that they had with the WVWRI, they wanted our data in 3RQ. But we did not receive a mini-grant. We applied but we didn't get it. We actually never even got a letter telling us that we didn't receive funds, we just saw the announcement and we were not in it. By that time we had funded the project in other ways, so it wasn't a big deal.

While 3RQ's rejection was not a major loss for Mandy's organization, 3RQ's growing position of authority as a gatekeeper for watershed science in the region set an expectation that groups would want to partner with 3RQ, regardless of whether or not they had received a mini-grant. This was a disempowering experience for other groups with a long history of water monitoring. For example, Colcom would eventually request that Mandy's organization contribute their data to 3RQ's database, but their exclusion from 3RQ's official research program meant that they would have a reduced role in determining how 3RQ might use their data.

In other instances, groups applied for 3RQ mini-grants for the explicit purpose of leveraging their data. Rita Levitt, the director of a Pennsylvania-based watershed association noted, "I mean, it is good to have your own database, but at the same time, you know, a central repository where hopefully it will never disappear, that was, for us, the goal." Decisions to partner with 3RQ on terms of data management were echoed by others who joined the KI. Water monitoring groups needed technologies to store rapidly accumulating data, but what they really desired was assistance from 3RQ's watershed scientists that this arrangement implied. Many people like Rita believed that partnering with a respected research institution would bring legitimacy to their data and reveal hidden evidence of pollution they themselves could not see.

Lisa, Mandy, and Rita's data were stored in a database and GIS system called the QUEST Data Management Tool. While this tool can be evaluated as a technical component of 3RQ's KI, it also has social significance a it's architecture echoed 3RQ's partnership structure—data entering the system was "tiered" to distinguish its source. Within this classification scheme, bi-weekly samples gathered by 3RQ's research partners were assigned to Tier 1. Tier 2 was reserved for data generated by automated data logger stations. Data from grassroots monitoring programs were placed in Tier 3 (Figure 3). Tiered data made sense to 3RQ's development team, particularly when having to work with regulatory agencies. 3RQ's program director noted:

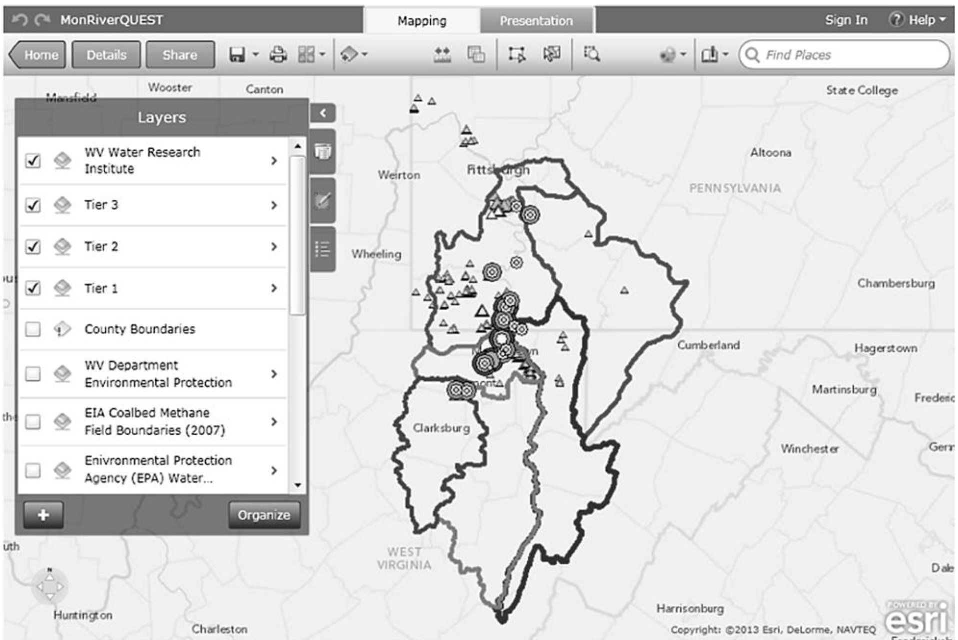

Figure 3. A screen capture of 3RQ's QUEST Data Management Tool (circa 2013)[1]

> We actually had conversations with the EPA and with the different state agencies, Pennsylvania DEP [Department of Environmental Protection] and West Virginia DEP, early on during the brainstorming phases of bringing on the volunteer component into the program, and that was one of the things that was identified. Whenever they are looking at the data from our website, they wanted to be able to distinguish between what we are collecting and what volunteers are collecting. And then, further, which volunteers are collecting—how much confidence can we give in this data that is being provided.

The tier system was born out of a need to define the characteristics and quality of 3RQ's data. However, representatives from smaller watershed groups I met with argued that 3RQ's data scheme also reflected the KIs overall political landscape. 3RQ's leaders made demands for their data in order to conduct scientific research, but diminished the importance of using data to advocate for impacted communities. These concerns were made plain in my conversation with Lisa Greenfield:

> What I don't see QUEST doing at this time, at least not in the way that maybe I would like to see it done, is then turning around, taking this data, and being the leaders—telling our elected officials that this is happening to our rivers and streams and this is what we need to do to protect them. That gets back into the big questions about the Ivory Tower, and who funds your research. I have opinions about the motivations behind some of this research. We might have all this data on our watershed, but how is that helping improve water quality broadly across the state and across the region? Yeah, we hope that nothing bad ever happens, but if it would, it wouldn't be the researchers marching down to Charleston [West Virginia's state capital], it would be us. I don't think they would help us.

When a number of watershed groups brought up the issue of 3RQ's reluctant support for advocacy at a regional meeting of mini-grantees, 3RQ's leaders countered that using data for research purposes could produce meaningful changes in environmental governance. They furthermore argued that, since 3RQ is part of West Virginia State's designated Water Research Institute, they were not in a position to use data beyond the purposes of research.

Renegotiating the Terms of Knowledge Infrastructures

Growing discord between how grassroots groups and research partners envisioned the purpose of the 3RQ threatened to unravel the KI. Numerous mini-grantees began to question their commitments to a KI that did not help them address their immediate environmental concerns. Similar complaints came from monitoring groups outside 3RQ that had been pressured to contribute data to the QUEST Data Management Tool.

These complaints had an interesting effect. By 2015, Colcom and WVWRI had invested more than $1.6 million to establish 3RQ as a regional hub for water monitoring research. 3RQ's leaders and funders took note of growing dissatisfaction and began to reevaluate the effectiveness of the infrastructure KI they had built. 3RQ modified the QUEST Data Tool tiers to indicate which protocols were used when collecting data, rather than what kind of organization did the collecting. Tier 3 now denotes data verified by an analytical lab, Tier 2 includes data collected with protocols such as ALLARM's, and Tier 1 is for data collected by individuals without known quality controls. Breaking the symbolic link between data's source and data's legitimacy was significant for nonprofessional groups who felt marginalized by 3RQ's expert-centric power structure.

A second major change came in June of 2015, when Colcom awarded 3RQ a fourth grant for $350,000. These funds established a program called REACH (Research Enhancing Awareness via Community Hydrology) and brought on four outreach coordinators to serve as links between 3RQ's researcher partners and local watershed groups (West Virginia University, 2015). REACH and changes to the tiered data structure represented a shift towards greater capacity for empowerment within 3RQ's KI.

Discussion

On their surfaces, the knowledge infrastructures designed by water monitoring networks in the Marcellus Shale are surprisingly similar—they propagated standardizes protocols, provided access to testing equipment, offered training to new members, developed a means to work with data, and created pathways to partner with scientific experts. These "internetworks of people, artifacts, and institutions" (Edwards et al., 2013: 23) addressed the needs of NY Water Sentinels' and 3RQ's affiliates and were constructed for similar reasons. People believed that investing in KIs would bring together diverse resources and knowledge to address shale oil and gas extraction's risks to watersheds. However, the participation models adopted in these infrastructures significantly impacted how stakeholders retained control in decision-making processes.

These two studies shed light on the nature of power sharing in KIs. In the case of the NY Water Sentinels, member groups enjoyed a high degree of autonomy to address new environmental pollution concerns as they arose. Their grassroots governing system afforded mechanisms for individuals to influence daily operations and ask new questions with their science. By comparison, local watershed organizations that aligned with 3RQ gained access to professional-grade resources and mini-grants brought forth new equipment and staff. However, 3RQ's partnership structure reinforced the authority of watershed experts while claiming to do co-designed research. Mini-grantees were able to hook into a sophisticated KI, but were immediately marginalized by the constraining priorities of 3RQ's research partners.

The two studies also demonstrate that the inner workings of KIs change over time when some stakeholders begin to assert greater influence. This was seen at two distinct points in the NY Water Sentinels' development. One occurred when a number of individuals inserted new objectives into their daily monitoring activities. The other turning point coincided with the NY Water Sentinels becoming a sub-program of the Sierra Club. For 3RQ, power shifts occurred for different reasons. Despite being part of one of the most resource-rich water monitoring networks in the Marcellus Shale, many of 3RQ's member organizations had little control in directing 3RQ's KI. Dissatisfaction became visible when members voiced concern about the ways their data was being managed and how research partners responded to their advocacy needs. Rather than breaking down, 3RQ's KI was transformed by tactics like choosing not to share data. The REACH initiative and changes to QUEST's tier structure represented a ceding of power; they illustrate how marginalized groups can alter KIs through various forms of resistance.

Finally, the NY Water Sentinels and 3RQ provide insights into the nature of empowerment in KIs. Corbett and Keller (2005a) offer a framework to assess empowerment and empowerment capacity at different scales: at the level of the individual and at the level of community. When brought to the study of KIs, this framework exposes some of the tradeoffs that occur when building KIs. Individuals who viewed landfill waste as a major threat were empowered by the NY Water Sentinels governing system, but one could also argue that the durability of their KI suffered due to internal frictions and competing objectives. Aligning with the Sierra Club may have saved the KI, but the constraints that come with this new partnership could, in the future, disempower the organizing capacities of affiliated monitoring groups. These are significant findings that deserve additional research into how KIs can effectively bridge dispersed research communities while maximizing capacity for collective empowerment.

Whether or not REACH will empower grassroots groups who invested in 3RQ's KI remains to be seen. It is likely that some mini-grantees will find some degree of empowerment by working with 3RQ's new community outreach coordinators; for instance, by having more resources to interpret their data. Generating long-term empowerment capacity to deal with environmental impacts is less certain. Academic researchers would have to share resources and utilize local knowledge in their work, thus yielding entrenched power to the voices and science of nonprofessionals.

Conclusion

Susan Leigh Star (1999: 382) once agued that "because infrastructure is big, layered, and complex, and because it means different things locally, it is never changed from above. Changes take time and negotiation, and adjustment with other aspects of the systems are involved. Nobody is really in charge of infrastructure." The arguments offered in this paper complement Star's sensibility. The Marcellus Shale water monitoring community emerged in order to deal with complex environmental and public health risks introduced by shale oil and gas extraction. Those who came together to build KIs for water monitoring research brought with them a wide spectrum of resources, expertise, and objectives. In studying this community, I have found it important to not only evaluate how KIs emerge, but also how power plays out in their emergence. What one finds is that KIs, even when seemingly stable in their leadership and intended purpose, are indeed dynamic spaces where relationships of power are rarely settled.

Subsequently, one must also give consideration to how KIs empower and disempower people in their daily lives. Many other regions in U.S. and abroad are paying close attention to public responses to oil and gas extraction's health and environmental threats in the Marcellus Shale. States with recently discovered shale formations, such as Florida, are setting regulatory frameworks that will determine how they assess the risks of hydraulic fracturing. Other states are shutting down channels of public participation and regulatory transparency. Wyoming recently criminalized citizen data collection on "open land"—land outside the jurisdiction of established cities and town (Pidot, 2015). In North Carolina, legislators outlawed disclosures of hydraulic fracturing chemicals in order to attract energy companies (Rosenberg, 2014).

Concerned citizens in these at-risk geographies are evaluating effective strategies for political resistance. Decisions made within the Marcellus Shale advocacy community will almost certainly propagate there and elsewhere. It is therefore critical to understand how these strategies—civil society science being one—struggle and succeed in overcoming barriers of public participation and influence. Knowledge infrastructures that emerge in these spaces can generate and curate new knowledge, is evident in many previous studies, but they can also assist marginalized communities to build capacity and mobilize resources when empowerment is a set intention in their design.

Note

1 The current version of the QUEST Data Management Tool can be found at http://3riversquest.org/ (Last accessed July 27, 2015).

Acknowledgements

This research was supported in part by the National Science Foundation, under Award #1331080 and Award #1126235. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation.

References

Bijker WE, Hughes T & Pinch T (eds) (1987) *The Social Construction of Technological Systems: New Directions in the Sociology and History of Technology*. Cambridge, MA: MIT Press.

Borgman CL, Wallis JC & Mayernik MS (2012) Who's Got the Data? Interdependencies in Science and Technology Collaborations. *Computer Supported Cooperative Work (CSCW)* 21(6): 485–523.

Bowker GC, Baker KS, Millerand F & Ribes D (2010) Toward information infrastructure studies: Ways of knowing in a networked environment. In: Hunsinger J, Klastrup L & Allen MM (eds) *International Handbook of Internet Research*. London, UK: Springer, 97–117.

Bowker GC & Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Corbett, JM & Keller CP (2005a) An Analytical Framework to Examine Empowerment Associated with Participatory Geographic Information Systems (PGIS). *Cartographica: The International Journal for Geographic Information and Geovisualization* 40(4): 91–102.

Corbett, JM & Keller CP (2005b) Empowerment and Participatory Geographic Information and Multimedia Systems: Observations from Two Communities in Indonesia. *Information Technologies and International Development 2*(2): 25–44.

Corburn J (2005) *Street Science: Community Knowledge and Environmental Health Justice*. Cambridge, MA: MIT Press.

Donlin P (2010) DEP investigating frack fluid spill near Hughesville. *Lock Haven Express*, 23 November. Available at: http://www.lockhaven.com/page/content.detail/id/527442.html (accessed 14.4.2015).

Donohue DM (2015) DEC extends Hyland landfill comment period. *Olean Times Herald*, 19 January. Available at: http://www.oleantimesherald.com/news/article_81aff232-a028-11e4-82a7-877ba728f765.html (accessed 27.7.2015).

Edwards PN, Bowker GC, Jackson SJ & Williams R (2009) Introduction: An agenda for infrastructure studies. *Journal of the Association of Information Systems* 10(5): 364 – 374.

Edwards PN, Jackson SJ, Chalmers MK, Bowker GC, Borgman CL, Ribes D, Burton M & Calvert S (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor, MI: Deep Blue.

Edwards PN, Mayernik MS, Batcheller AL, Bowker GC & Borgman CL (2011) Science Friction: Data, metadata, and collaboration. *Social Studies of Science* 41(5): 667–690.

Entrekin S, Evans-White M, Johnson B & Hagenbuch E (2011) Rapid Expansion of Natural Gas Development Poses a Threat to Surface Waters. *Frontiers in Ecology and the Environment* 9(9): 503–511.

Fischer F (2000) *Citizens, Experts, and the Environment: the Politics of Local Knowledge*. Durham, NC: Duke University Press.

Gouveia C, Fonseca A, Câmara A & Ferreira F (2004) Promoting the use of environmental data collected by concerned citizens through information and communication technologies. *Journal of Environmental Management* 71(2): 135–54.

Hardman E (2014) Is there leachate on tap? A report from Bath, NY. *Rochester Indymedia*, 22 September. Available at: http://rochester.indymedia.org/node/104041 (accessed 22.4.2015).

Hughes T (1987) The evolution of Large Technological Systems. In: Bijker WE, Hughes T & Pinch T (eds) *The social construction of technological systems: New directions in the sociology and history of technology*. Cambridge, MA: MIT Press, 51–82.

Irwin A (2001) Constructing the Scientific Citizen: Science and Democracy in the Biosciences. *Public Understanding of Science* 10(1): 1–18.

Jalbert K & Kinchy AJ (2015) Sense and Influence: Environmental Monitoring Tools and the Power of Citizen Science. *Journal of Environmental Policy & Planning* 2: 0–19.

Jalbert K, Kinchy AJ & Perry SL (2014) Civil Society Research and Marcellus Shale Natural Gas Development: Results of a Survey of Volunteer Water Monitoring Organizations. *Journal of Environmental Studies and Sciences* 4(1): 78–86.

Kinchy AJ & Perry SL (2012) Can Volunteers Pick Up the Slack? Efforts to Remedy Knowledge Gaps About the Watershed Impacts of Marcellus Shale Gas Development. *Duke Environmental Law & Policy Forum* 22(2): 303–339.

Lave R (2012) Neoliberalism and the Production of Environmental Knowledge. *Environment and Society: Advances in Research* 3(1): 19–38.

Llewellyn GT, Dorman F, Westland JL, Yoxtheimer D, Grieve P, Sowers T, Humston-Fulmer E & Brantley SL (2015) Evaluating a groundwater supply contamination incident attributed to Marcellus Shale gas development. *Proceedings of the National Academy of Sciences* 112(20): 201420279.

Malone S, Kelso M, Auch T, Edelstein K, Ferrar K & Jalbert K (2015) Data inconsistencies from states with unconventional oil and gas activity. *Journal of Environmental Health and Science* 50(5): 37–41.

Mantius P (2013) New York Imports Pennsylvania's Radioactive Fracking Waste Despite Falsified Water Tests. *DC Bureau Natural Resources News Service*, 14 August. Available at: http://www.dcbureau.org/201308148881/natural-resources-news-service/new-york-imports-pennsylvanias-radioactive-fracking-waste-despite-falsified-water-tests.html (accessed 27.7.2015).

Moore K (2006) Powered by the People: scientific authority in participatory science. In: Frickel S & Moore K (eds) *The New Political Sociology of Science: Institutions, Networks, and Power*. Madison, WI: University of Wisconsin Press, 299–323.

Moran E (2015) *License to Dump: Despite Ban, New York Permits Pennsylvania to Dump Radioactive Fracking Waste Inside Our Borders*. Environmental Advocates of New York. Available at: http://www.eany.org/our-work/reports/license-dump-february-2015 (accessed 27.7.2015).

Morello-Frosch R, Pastor M, Sadd JL, Porras C & Prichard M (2005) Citizens, science, and data judo: Leveraging secondary data analysis to build a community-academic collaborative for environmental justice in Southern California. In: Schulz AJ & Parker EA (eds) *Methods in community-based participatory research for health*. San Francisco, CA: Jossey-Bass, 371–392.

NYS Department of Environmental Conservation (2006) *Regulations Part 371: Identification and Listing Of Hazardous Wastes*. Available at: http://www.dec.ny.gov/regs/14897.html (accessed 4.22.2015).

O'Rourke D & Macey GP (2003) Community environmental policing: Assessing new strategies of public participation in environmental regulation. *Journal of Policy Analysis and Management* 22(3): 383–414.

Ottinger G (2009) Buckets of Resistance: Standards and the Effectiveness of Citizen Science. *Science, Technology & Human Values* 35(2): 244–270. http://doi.org/10.1177/0162243909337121

Overdevest C & Mayer B (2008) Harnessing the Power of Information through Community Monitoring: Insights from Social Science. *Texas Law Review* 86(1493): 1493–1526.

Pfeffer, Max J, and Linda P. Wagenet (2007) Volunteer environmental monitoring, knowledge creation and citizen–scientist interaction. *The SAGE handbook of environment and society. Sage, London*, 235-250.

Pidot J (2015). Wyoming law against data collection: Protecting ranchers by ignoring the environment. *Slate.com*, 11 May. Available at: http://www.slate.com/articles/health_and_science/science/2015/05/wyoming_law_against_data_collection_protecting_ranchers_by_ignoring_the.html (accessed 12.2.2016).

Pinch T & Bijker WE (1987) The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In: Bijker W, Hughes T & Pinch T (eds) *The Social Construction of Technological Systems*. Cambridge, MA: The MIT Press.

Puko T (2013) Radioactive fracking debris triggers worries at dump sites. *Pittsburgh Tribune-Review*, 11 May. Available at: http://triblive.com/business/headlines/3945499-74/gas-radiation-radioactivity#axzz3aX2cieYp (accessed 27.7.2015).

Ribes D & Finholt TA (2009) The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10(5): 375–398.

Rosenberg M (2014) North Carolina Senate outlaws disclosure of fracking fluid secrets. *Thomson Reuters*, 22 May. Available at:http://www.reuters.com/article/us-usa-fracking-secrets-idUSBREA4L0YC20140522 (accessed 12.2.2016).

Savan B, Morgan AJ & Gore C (2003) Volunteer Environmental Monitoring and the Role of the Universities: The Case of Citizens' Environment Watch. *Environmental Management* 31(5): 561–568.

Shirk JL, Ballard HL, Wilderman CC, Phillips T, Wiggins A, Jordan R, McCallie E, Minarchek M, Lewenstein BV, Krasny ME & Bonney R (2012) Public Participation in Scientific Research: a Framework for Deliberate Design. *Ecology and Society* 17(2): 29.

Soraghan M (2011) Frack Study's Safety Findings Exaggerated, Bush EPA Official Says. *The New York Times*, 20 May. Available at: http://www.nytimes.com/gwire/2011/05/20/20greenwire-frack-studys-safety-findings-exaggerated-bush-65374.html (accessed 14.4.2015).

Star SL (1999) The Ethnography of Infrastructure. *American Behavioral Scientist* 43(3):377–391.

Star SL & Ruhleder K (1996) Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research* 7(1): 111–134.

US Energy Information Agency (1993) *Drilling Sideways - A Review of Horizontal Well Technology and Its Domestic Application, DOE/EIA-TR-0565*. Available at: http://www.eia.gov/pub/oil_gas/natural_gas/analysis_publications/drilling_sideways_well_technology/pdf/tr0565.pdf (accessed 28.7.2015).

US Energy Information Agency (2013) *Technically Recoverable Shale Oil and Shale Gas Resources: An Assessment of 137 Shale Formations in 41 Countries Outside the United States*. Available at: https://www.eia.gov/analysis/studies/worldshalegas/archive/2013/pdf/fullreport_2013.pdf (accessed 10.2.2016).

Wagenet LP & Pfeffer MJ (2007) Organizing Citizen Engagement for Democratic Environmental Planning. *Society & Natural Resources*, 20(9): 801–813.

West Virginia University (2013) Regional water quality monitoring program receives $508,000 grant; allows researchers to identify long-term trends. *WVU Today*, 24 September. Available at: http://wvutoday.wvu.edu/n/2013/09/24/regional-water-quality-monitoring-program-receives-508-000-grant-allows-researchers-to-identify-long-term-trends (accessed 22.4.2015).

West Virginia University (2015) $350,000 grant expands WVU water quality monitoring of three rivers. *WVU Today*, 3 June. Available at: http://wvutoday.wvu.edu/n/2015/06/03/350-000-grant-expands-wvu-water-quality-monitoring-of-three-rivers-community-outreach-funded-for-targeting-trouble-spots (accessed 27.7.2015).

Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases

Éric Dagiral

Université Paris Descartes, CERLIS, Sorbonne Paris Cité, France / eric.dagiral@parisdescartes.fr

Ashveen Peerbaye

LISIS, Université Paris-Est, France

Abstract

This paper provides an ethnographical study of the ways in which infrastructure matters in the production of knowledge in the social worlds of rare diseases. We analyse the role played by a relational database in this respect, which exists at the crossroads of a large and complex network of individuals, institutions, and practices. This database forms part of a "boundary infrastructure", in which knowledge production constitutes one output of infrastructural work, that needs to be articulated with other kinds of activities and matters of concern. We analyse how members of the network negotiate the place and forms of knowledge production in relation to these other purposes, and highlight the political nature of the distinction between knowledge and information, which frames collective action. We also show how infrastructural inversion serves to articulate knowledge production with other forms of mobilisation, thereby shaping and reconfiguring the boundary infrastructure as a whole.

Keywords: knowledge infrastructures, boundary infrastructures, relational databases, rare diseases

Introduction

This paper provides a conceptualisation of the role played by infrastructure in the production of knowledge on human rare diseases. Despite the rarity of individual cases, about 30 million persons living in Europe today are estimated to suffer from one of these disorders, of which over 6000 have been identified, with many being early-onset, chronic, degenerative and invalidating conditions. Individuals and organisations exposed to rare diseases (e.g. patients and their relatives, physicians, scientists, healthcare professionals) have to face situations in which the lack of knowledge heavily hampers diagnosis, referrals to specialists, medical care provision, as well as clinical and therapeutic research. Over the last three decades, scientific research centres, medical and healthcare organisations, pharmaceutical industrialists, public health institutions and patient advocacy groups have gathered into a heterogeneous network devoted to the cause of rare diseases (Huyard, 2012). In France and elsewhere in Europe these various actors have gradually begun to cooperate (frequently without consensus) in order to gather resources, align political agendas, and manage a large and ever-extending variety of projects related to rare diseases. Some of these projects aim to foster biomedical research in order to document epidemiological, clinical and genetic aspects of understudied diseases and disorders. Other projects strive to gather and distribute scarce information to healthcare organisations, professionals, or the general public. Still others work for the promotion of orphan drug development by the pharmaceutical industry, or the empowerment of patients and their relatives, for instance through the building of online communities, participation in clinical trials, and struggle for equal access to available treatment and healthcare. These patterns of cooperation across the network have come to rely ever more heavily upon an array of work routines and tools, such as online databases, diagnostic expert systems, classification systems, indexes, registries and so on. This set of distributed technical, informational and organisational resources constitutes an infrastructure that now shapes ways of knowing, working and living with rare diseases.

Recent scholarship in STS adopts a broad definition of infrastructure and highlights their crucial role in the production of knowledge. Bowker et al. (2010: 98) suggest envisioning knowledge infrastructures as "pervasive enabling resources in network form" that allow "knowledge work" to be performed. This definition departs from the conventional representation of infrastructure as a mere machinery of "tubes and wires", to include a wide range of technologies and organisations that span large-scale sites and instruments devoted to scientific research (e.g. "supercolliders, orbiting telescopes, supercomputer centres, polar research stations, national laboratories"), institutional and technological structures that buttress the functioning of science ("funding agencies, professional societies, libraries and databases, scientific publishing houses, review systems, and so on"), as well as the various users, mediators and professionals that are involved in the design, implementation, and management of shared digital services and resources ("data and code repositories, best practices and standards development, visualisation tools and high performance computing, and so on"). This definition is very illuminating since it provides a framework for analysing how infrastructure reconfigures work practices, and especially scientific work. Biomedical research and the life sciences have provided major fields of inquiry in this regard. The importance of genome databases in the "the canonical scientific act for our times (sequencing the genome)" (Bowker, 2005: 30) is for instance well documented. Studies of knowledge production regimes in the sciences highlight the increased reliance upon shared facilities and instrumentation, online digital databases, as well as standards for the publication and sharing of data and metadata (Millerand & Bowker, 2009). These transformations are predicated on technologies, capacities and practices for storing, analysing, representing and circulating information. They modify the speed and scale at which these operations and exchanges take place. As Bowker and Star (1999: 108) argue "infrastructure does more than make work easier, faster or more efficient; it changes the very nature of what is understood by work." In the case of knowledge production, these transformations introduce novel forms of publication and validation of scientific results (Hilgartner, 1995), affect patterns of scientific collaboration (Parker et al., 2010) and ways of knowing (Strasser, 2011): they change the very nature of what is understood by knowledge.

A growing body of literature has started to take into account such infrastructural transformations – very often gathered under such labels as e-science, cyberscience or cyberinfrastructure – in a large variety of disciplines, ranging from life sciences to ecology, biodiversity, earth and climate sciences, and the humanities (Miller & Edwards, 2001; Hine, 2006b; Olson et al., 2008; Dutton & Jeffreys, 2010). Social studies of databases are of particular importance in this respect. Bowker (2000) shows that the convergence of heterogeneous databases in biodiversity sciences raises the issue of how various social and political values might be embedded into the emerging infrastructure, and be made to coexist. Examining the development of a mouse genome mapping resource, Hine (2006a) argues that databases are more and more often configured and used as genuine research tools in their own right. Their mobilisation in the laboratory introduces additional mediations and challenges in the organisation and outcomes of knowledge work. The negotiation of rules for data entry, for instance, reveals the complex reorderings that such resources entail, and casts new light on the issue of cooperation in molecular biology as analysed by Knorr Cetina (1999: 234-240). Millerand (2011) in her study of a large-scale database devoted to arctic research, gathering geophysical, biological, medical and sociological data, pinpoints the differences in instrumental practices and the various significations attributed to data, and explores their impact on the division of scientific labor across disciplines and epistemic cultures. Leonelli provides an understanding of the ways in which "the worlds of data infrastructures and knowledge production inform each other" (Leonelli, 2013: 513), by foregrounding the role of data-sharing resources, such as databases, in the production of scientific evidence in contemporary biology. Her definition of data as material artifacts which are mobilised in relation to specific contexts of knowledge production captures the different modalities through which data integration is performed to produce new knowledge, according to a variety of goals, methods, strategies and norms (Leonelli, 2015).

Our own contribution aims to similarly emphasise the crucial ways in which infrastructural issues come to matter in the production of knowledge in the social worlds of rare diseases. Of particular importance here is the role played by a relational database devoted to rare diseases and orphan drugs, whose setting up, maintenance and use can be seen as a major achievement of the collectives involved. However, this database will be shown to exist at the crossroads of a large and complex network of individuals, institutions, and practices, in which the basis of collective mobilisation is quite blurry, and not centred at the outset on knowledge production – though undoubtedly the circulation and use of knowledge and information are crucial issues here for collective action. In contrast to situations where the production of knowledge, and especially scientific knowledge, is the core legitimate focus and outcome of the organisation of work (e.g. in laboratories, scientific collaborative networks, disciplinary or transdisciplinary research communities, etc.), our case provides a context in which the involvement of different communities extends well beyond this goal. The database forms part of a larger infrastructure, in which knowledge production constitutes one output of infrastructural work, that needs to be furthermore articulated with other kinds of activities and matters of concern. Bowker and Star (1999: 313) resort to the concept of "boundary infrastructure" to acknowledge these situations where "[an] infrastructure serves multiple communities of practice simultaneously, be these within a single organization or distributed across multiple organizations". They argue that "[what] we gain with the concept of boundary infrastructure over the more traditional unitary vision of infrastructures is the explicit recognition of the differing constitution of information objects within the diverse communities of practice that share a given infrastructure" (Bowker & Star, 1999: 314). Our study takes up this concept and explores its practical and theoretical implications, by examining how members of the network negotiate the place and forms of knowledge production in relation to the other purposes they pursue in various collaborative projects. This leads us to put forward two main claims. The first has to do with the political nature of the distinction between knowledge and "mere" information, which lies at the very heart of many debates between actors. Here, this kind of boundary work (Gieryn, 1983) serves as a resource in strategies to embed competing visions and goals into the boundary infrastructure, define priorities, and allocate resources for carrying out different tasks related to rare disease initiatives. The second is that in order to negotiate the political, moral and epistemic dimensions of the boundary infrastructure they contribute to and rely on, actors resort to infrastructural inversion (Bowker, 1994): they discuss explicitly of the infrastructure itself, and strive to represent its inner workings, shortcomings and desirable evolutions. Infrastructural inversion, therefore, is not only a methodological lens for the analyst to capture how things like databases and classification systems are embedded in the many practices of collectives engaged in the field of rare diseases. It is also constitutive of the practices of these collectives themselves.

The arguments presented in this paper are grounded in extensive fieldwork carried out between 2007 and 2013, to explore the social worlds of rare diseases at the French and European levels. The authors jointly conducted a four-year ethnographic study of the "Rare Diseases Platform", a European-level entity created in the early 2000s, and located in Paris (France). The Platform itself is comprises of six distinct organisations: a data-based resource centre (Orphanet) belonging to the French Institute of Health and Medical Research; a help line devoted to providing support and information on rare diseases to the general public (Maladies Rares Info Services); a consortium for the funding and promotion of research and healthcare activities in the field of rare diseases (GIS-Institute for Rare Diseases); two patient-driven federations (Alliance Maladies Rares in France, and EURORDIS at the European level); and the French Muscular Dystrophy Association (AFM-Téléthon) which played a major part in gathering those organisations into a "platform" in the first place. During that period, we collected data from regular participant observation of individual and collective everyday work activities at Orphanet and MRIS (situations involving database manipulation, editorial activities, discussions at project meetings, maintenance work). We analysed various documents and work materials, including technical worksheets, reference documentation, meeting and activity reports, internal memos, and transcripts from electronic communications. We also regularly attended collective events in which the different members of the Platform gathered as a whole (board and committee meetings, scientific and associative events, annual March for Rare Diseases, etc.) In addition to everyday informal interactions, we conducted fifty semi-structured interviews with twenty-six team members and managing directors from these six organisations, focusing on aspects related to scientific and extra-scientific collaboration, information gathering and exchange, and involvement in technological projects pertaining to infrastructure. Follow-up materials were gathered in 2011-2013, through thirteen interviews with database managers and technicians, and experts in health information systems involved in many of the Platform's projects. These focused mainly on the integration of novel web ontologies and standards into the Platform's existing infrastructure.

We begin by examining how the boundary between knowledge and information is negotiated inside the European Rare Diseases Platform and contributes to frame collective action. We then show that what counts *as* knowledge infrastructure and what counts *in* a knowledge infrastructure are materially enacted within a relational database. Finally, we move on beyond the database itself to reveal how infrastructural inversion serves to articulate knowledge production with other forms of mobilisation, thereby shaping and reconfiguring the boundary infrastructure as a whole.

Negotiating the Boundary Between Knowledge and Information in the European Rare Diseases Platform

Knowing What Counts as Knowledge

Issues related to knowledge production have been from the very outset a central concern for the different organisations that gathered together in the early 2000s to form the Rare Diseases Platform. The main challenge in this respect has been to articulate a framework taking into account a broad understanding of what constitutes "knowledge" in the social worlds of rare diseases, as well as securing and allocating resources for its production and mobilisation in concerted action. One aspect of this problem refers to the necessity of addressing the lack of biomedical knowledge on understudied low-prevalence diseases and disorders, by fostering scientific, discipline-based endeavours, in clinical research, experimental medicine and biology, epidemiology and so on. As one director of the French Muscular Dystrophy Association put it to us:

> "In a field such as rare diseases, boundaries are more blurry than with common diseases. This is problematic because EU-funded research projects, for example, do not have the same rules for funding whether one is at the level of the Department of health or the Department of research. But, in the case of rare diseases you do not have health teams on side, and scientific research teams on the other. In the teams, you find both the research part mixed with the health part. In many cases, the clinician provides medical care, does fundamental research, is involved in information activities, etc.

> This integrated model corresponds to the way we wanted to function as a platform. The idea is that in order to make progress on such a topic, we need to advance in a systematic, global fashion, tackling things from all sides, provided that there is a minimum of coordination and consistency in our approaches."

However, construing knowledge as scientific knowledge alone was not deemed satisfactory to tackle the social problem posed by rare diseases. As the Platform progressively gained organisational and institutional reality, discussions between participants about knowledge issues and how to address them started to be framed through the broader motto of the "fight against ignorance":

> "Managing to spell out these problems of social ignorance through the category of rare diseases was far from self-evident. It is true that genetics and science have contributed to build a collective vision of rare diseases. However, when the Platform was created, the actors' perception was that beyond the scientific aspect of [rare] diseases, we had to struggle in favour of populations whose existence was being denied from a medical, scientific but also civic standpoint, and who weren't being listened to. [...] Our goal was to build acceptance about the fact that this small minority of people and situations were included inside society, had rights as everybody else, and that there was a need for specific modalities to secure their exercise of these rights. [...] We had a strong commitment to bring about societal change, by using all the levers of modern information and communication." (*French Muscular Dystrophy Association director*)

In their discussions and negotiations to find common ground for collective action, "information" rather than "knowledge" quickly became the category of choice under which participants of the Platform framed their involvement and publicised their "fight against ignorance" of rare diseases. One of the main objectives of the Platform's programme officially became to "develop information on rare diseases for patients, health professionals and the general public", a priority which was integrated as such into national plans and strategies for rare diseases (notably the French National Plan on Rare Diseases in 2005, but also in Bulgaria, Portugal and Spain), as well as into the policy framework for rare diseases defined by the European Union Committee of Experts on Rare Diseases (Aymé & Rodwell, 2013). For all parties involved, including the Platform's main funding institutions (AFM-Téléthon and the French Ministry of Health), resorting to the category of information better allowed to take into account the fact that their commitment exceeded the production of novel biomedical knowledge, and included knowledge-related activities that were often not sanctioned as such by institutions: providing information about diseases for instance, or fostering relationships between hospitals, laboratories, companies and patients. This focus on information brought to the fore practical issues related to activities of gathering, consolidating, formatting, connecting and circulating heterogeneous resources across various communities. It also fostered a more "representational" vision of knowledge on rare diseases, understood as something that must not only be produced, but also be made visible and mediated in order for rare diseases to be recognised as a social problem in its multiple dimensions (Dagiral & Peerbaye, 2012).

The trajectory of Orphanet, one of the Platform's main components devoted to information development, clearly exemplifies this strategy. The existence of Orphanet as an organisational entity predates its inclusion in the Platform. It was created in 1997 as a unit of the French National Institute of Health and Medical Research (Inserm), resulting from longstanding efforts (going back to the 1970s and 1980s) from the part of a handful of physicians and geneticists who were being confronted with diagnosis problems in cases encountered during consultations, for which documentation in the medical literature was very scarce or unavailable. These specialists strived to gather expertise, monitor the scientific literature and create classifications of these rare symptoms in a computerised form. One aspect of the creation of Orphanet therefore results from a clinical and scientific concern that at the same time encompassed a technological issue: creating from scratch a computer database that could be queried as an expert system to help establish diagnoses in infrequently encountered situations. Of course, as many authors studying infrastructures have pointed out, an infrastructure rarely ever springs *ex nihilo*, but rests upon an "installed base" of pre-existing elements, from which it inherits its strengths and limitations (Star & Ruhleder, 1996: 113). In this case, these elements included paper documents piled up in bulky folders, punch cards, old computer programming languages such as Fortran, videodiscs, CD-ROMs and now-obsolete software applications (Dagiral & Peerbaye, 2013). During the 1990s, and even more in the early 2000s, following its inclusion in the Rare Diseases Platform, Orphanet moved from a physician-oriented diagnostic expert system to a multi-purpose instrument central to the Platform's information development strategy. This materialised in particular through the setting up of a public web portal for rare diseases and orphan drugs, named Orphanet[1], whose operation relied on a vast computerised relational database. Clearly, the system wasn't intended and designed at the outset to become the cornerstone of an overarching and ever-growing infrastructure supporting the activities of various organisations working in the field of rare diseases. It provided, however, the technological core around which different kinds of resources on rare diseases started to coalesce, in the guise of an online relational database warehousing very heterogeneous elements, to be articulated and used by multiple publics, in varied situations.

Orphanet is led today by a consortium of around 40 countries, coordinated by the French Inserm team. Teams in Europe and other parts of the world are responsible for gathering information on expert centres, medical laboratories, ongoing research and patient organisations in the field of rare diseases in their country, while the French coordinating team is responsible for the management of tools, standards and quality control procedures, but also provides rare disease inventories and classifications, and produces a rare diseases encyclopaedia. The "public" side of this initiative consists of the web portal, available in seven languages, which offers several types of resources aimed at patients and their families, patient organisations, as well as professionals – physicians, researchers, industry actors, and public health authorities. Starting with the definition of a disease and its clinical signs, the portal collates data related notably to epidemiological and genetic aspects, research projects, scientific publications, expert centres, diagnostic tests, clinical trials, and patient organisations associated with the given disease. In the early 2010s, this amounted to more than 6000 diseases or disorders, that existed on the "private" side as digital entities inside a relational database, on which different professionals intervened in order to update information, add newly identified diseases and resources, while also rethinking how all this information should be classified and interconnected, both from a cognitive and a technological perspective.

Orphanet's trajectory shares many characteristics with contemporary situations that may be encountered in a number of professional spheres. The digitisation of networks, the prevalence of databases, and the use of the Internet in everyday work inside or across organisations are common features that have introduced important reconfigurations in the distribution and collectivisation of tasks. They have brought a greater division of labour, changed how activities are measured and monitored, and increased the complexity of relationships between members of an organisation, as well as between organisations in cooperative work contexts.

The Platform as Boundary Infrastructure

It is tempting in our case to consider the Orphanet database and its interfaces (one of which is the web portal) as encapsulating the entire knowledge infrastructure of the Platform. This however would simply not be true. While the database was undeniably configured to function as a quasi-obligatory passage point for producing information and representing knowledge on rare diseases, it does not constitute in and of itself the whole infrastructure. This is due to the fact that, as Star and Ruhleder (1996) repeatedly remind us, infrastructure is a fundamentally relational phenomenon, not just something that "sits there": what is the daily work of one person is the infrastructure of another, what counts as an enabling infrastructure in one situation can become an obstacle in another. As a consequence, the perimeter of a knowledge infrastructure is not defined by its technical manifestation as a thing (a database, an online web portal, etc.) but by the shifting sociotechnical forms and organised practices that happen in situations of use across multiple communities of practice.

For an infrastructure to exist and function properly, investments are needed. Here, Laurent Thévenot's (1984) concept of "investment in form" proves analytically instrumental. As Keating and Cambrosio (2003: 38) have argued in the case of the establishment of biomedical platforms, the "reallocation of personnel and material entail[s] an investment in form that enable[s] previously heterogeneous equipment (in the largest sense of the term) to function together as a new sociotechnical unit". Thévenot (1984: 5-9) defines investment in form as "a costly operation to establish a stable relation with a certain lifespan", which can be performed through a "great variety of formatting operations, from the material constraint of standardization to the moral imperative of engagement, and the obligation of conventions":

> "Conforming and informing both require and are preceded by acts of giving form. This is why an 'investment in form', which might rely on different 'formats of information' [...] is the keystone that joins 'regulation' and 'objectivity'. The returns on such an investment, in terms of coordination, vary according to three dimensions: the temporal and spatial validity of the form, and the solidity of the material equipment involved. Once an investment has been made, it will have a 'temporal validity': that is, the period of time in which it is operative in a community of users. It will also have a 'spatial validity', which refers to the boundaries demarcating the community within which the form will be valid. This is why participating in the process of form-giving can be a means to prevent a standard from becoming external to one's own concerns, and therefore, potentially exclusionary." (Thévenot, 2009: 794)

Thévenot (2009) provides an insightful framework, which ties together nicely many of the themes we try to stress in this study: 1) a sense of collective commitment that rests upon and allows for the possibility of coordination; 2) a world-building relationship between persons and organisations that relies on the production of categories of likeness, equivalence and homogeneity through specific relations to things and their transformation; 3) a relation between "invested forms" and the engagement in specific modes of coordinated action they entail, which come to be deemed more effective, legitimate, desirable, and binding; 4) a certain disregard for the ordinarily assumed distinctions between cognitive, informational, technical and regulatory operations – understood as all partaking in the act of "giving form"; and 5) a focus on the characteristics of sustainability and the modalities of extension of the forms implemented.

A very relevant illustration of this process is provided by the use of the category of "platform" by members of the rare diseases community, in order to reflexively designate their engagement and modes of coordinated action. Indeed, this notion articulates a political commitment to a set of ideas and values with technical considerations, as well as spatial aspects (working together in connected spaces, whilst maintaining separate organisations and institutional arrangements):

> "[In calling ourselves a platform] we were probably influenced in part by the logic of technological research. [...] At the same time, alongside building and managing collective logistics and sharing technological resources, there was a strong will to appear as one structure. What is important in appearing as a platform is showing how we believe things should be approached: globally, systematically, mobilising actors with major levels of responsibility and commitment." (*French Muscular Dystrophy Association director*)

This meaningful inscription in a shared physical space of distinct forms of expertise and matters of concern also rested ultimately on operations of investment in form as defined above. This is a fact most actors were reminded of on an almost daily basis during our fieldwork, since the Platform was in the process of being physically relocated from one building to another in a Parisian hospital. Tense negotiations ensued about the distribution not only of working space, but also of workload and responsibilities among the platform's various entities. At their very heart lay the highly practical modalities of collaboration needed to ensure that knowledge work could be productively mobilised to further other kinds of activities, such as providing information and counselling, putting patients in touch with various organisations, or support the advocacy of rare disease associations.

Bowker and Star (1999) propose the concept of “boundary infrastructure” to capture how any working infrastructure, far from constituting a monolithic unit, provides an evolving system of boundary objects which different communities of practice (distributed within and across organisations) can simultaneously “plug into” in order to collaborate, all the while maintaining just enough local variation and just enough global consistency across sites for various kinds of commitments to work in concert (Star & Griesemer, 1989; Star, 2010). Envisioning the Rare Diseases Platform as an instantiation of such a boundary infrastructure is illuminating, since it puts to the fore the multiple, and sometimes competing, needs and visions that must be taken into account in order to “build and maintain productive relationships among people, organizations, and technologies” (Bietz et al., 2010: 245) devoted to rare diseases. Moreover, this allows a conceptualisation in which the material artifacts that constitute a regime of boundary objects – such as databases, expert systems or classification tools – are seen not only as collaboration-enabling mediators that can circulate between multiple communities, but also as resources that “can serve to *establish and destabilize* protocols themselves [and be] used to push boundaries rather than merely sailing across them”, as Lee (2007: 308) powerfully suggests in her study of a multidisciplinary collaborative design of a museum exhibition.

In the next section, we use this understanding of the Rare Diseases Platform as a boundary infrastructure to explore how knowledge activities permeate the daily work of a large number of its members. We strive to show the active, generative characteristic that reveals itself in the infrastructural work performed by the involved parties. Star and Ruhleder (1996: 114) answered the question: “When is infrastructure?” by stating that “an infrastructure occurs when the tension between local and global is resolved”. Our own answer aims at suggesting that, more often than not, a knowledge infrastructure occurs when the tension between local and global is reflexively and productively maintained.

Inside the Database: the Material Embeddedness of Knowledge

Knowledge (at) Work

In her social history of rare diseases, Caroline Huyard (2009: 475) insists on the fact that “the category of rare diseases was created with the very intention of restoring collaborative relationships between stakeholders who were unable to find common ground”. Her work approaches rare diseases as a boundary object that came about in order to create mobilisation around a cause that is first and foremost analysed as “political” (Huyard, 2012). In our own study, we try to supplement this view, by focusing on the concrete, everyday work of heterogeneous engineering (Law, 1987) that is needed to sustain the category as a boundary object and keep it afloat. This leads us firstly to highlight the important role played by various forms of knowledge activities that are centrally organised around Orphanet’s relational database, by focusing on the work of so-called “information scientists” (*documentalistes scientifiques* in French), who for one reason or another have to intervene in the “inner” workings of the database on a day-to-day basis. We then give an illustration of the way elements of the database are mobilised by other organisations in the Platform, focusing on how counsellors on the MRIS help line interact with Orphanet.

Information scientists at Orphanet are responsible for gathering information to “feed” the database, both by creating new entries and updating existing information. They can be found working individually, or in small teams, delving into external sources to find new or additional information about a disease, may it be through scientific publications, genetic databases, or websites listing experts, research laboratories, diagnostic centres and so on. They spend a good deal of time browsing online scientific databases such as PubMed, or querying Google, Google Scholar, and specialised search engines. The computer screens behind which they are busy browsing, reading, writing or talking to each other, typically display multiple windows, one of which might be open on the standard thesaurus of medical vocabulary (MeSH), another on an online catalogue of human genes and genetic disorders (OMIM), and occasionally a third window displaying a Wikipedia entry. These search activities allow the information scientists to complete and update fields in the Orphanet database through a dedicated in-house software interface, but also provides an opportunity to identify and take note of new diseases and resources that could be integrated in the future. In carrying out these tasks, Orphanet's information scientists are not simply looking for missing information through entirely routinised procedures, knowing in advance what they need to find and where. Rather, they are actively engaged in reading and interpretive activities, as well as operations to include and associate data from heterogeneous sources, which were not linked prior to their involvement. We see them for instance adding new types of materials to the database, such as hypertext links or detailed queries that will allow a user to find a collection of relevant scientific articles related to a specific disease within PubMed. This work requires familiarity with "database culture" – most often acquired on the job – and it is not limited to Orphanet's technical infrastructure, but extends more generally to mastering all existing specialised information sources in the field. Another common activity entailed by the need to "gather information" consists in identifying and interacting with various experts in order to enrol them as partners and help collect more information about a given disease. The information scientists thus spend a great deal of time trying to obtain data and requesting updates, by asking people to fill out forms, or appealing to European partners who can provide information through a shared online tool.

Information scientists at Orphanet also write – and get others to write – documents that will end up as electronic resources embedded in the texture of the database and the web portal. This concerns for instance "summaries", which form the centrepiece around which each disease entry and its related resources are arranged inside the database. The information scientists initiate first drafts, correct, edit, proofread, and distribute texts to experts and patient organisations in order to come up with, for example, articles for the general public or recommendations regarding "emergency procedures" for a given disease. One information scientist describes part of this process:

> "We send these documents to associations and experts at the same time. Everything that is related to medical aspects is treated by the experts. But we have an item that is named 'Additional emergency and hospitalisation guidelines', which are meant to be handed by emergency physicians to teams receiving patients for hospitalisation. These are things that a medical expert will typically never think or write about, very simple and commonsensical things, like adapting the size of the bed for a patient with Marfan syndrome who measures two meters, you see? These are also things one must think about, and we ask patient associations to pay attention to these details and add them to the documents."

Finding the "right" expert – a specialist, one gathers, who is at the same time renowned in her field, well versed in the clinical aspects of rare diseases, and willing to devote time to write and sign a text for Orphanet – is a challenge that proved for a long time a major hurdle for the organisation. In 2006, to mitigate the difficulty of getting experts involved in the writing process, Orphanet created a peer-reviewed scientific journal called the *Orphanet Journal of Rare Diseases*. From then on, this offered researchers an opportunity to publish their findings according to the canons of science, with the added advantage of indexation in the main databases (PubMed, Medline, ISI Web of Science, Google Scholar, Scopus, etc.) used by the information scientists in their work. The management of such distributed writing, which requires keeping track of the different versions and deadlines for a given text, relies on a host of memory practices (Bowker, 2005) and produces in turn an accumulation of various material traces: Excel spreadsheets, archives of email exchanges, printouts and so on. During this process of identifying, enrolling and assisting medical experts and patient associations in the production of texts, important choices are made that restate epistemic and political orientations and commitments, and affect what is present inside the database, and how this is presented in the various outputs.

Maladies Rares Info Services (MRIS) is another organisation part of the Platform that delivers information and provides support to people calling the help line with queries related to (potentially) rare diseases. We are immediately reminded of the importance Orphanet plays in these counsellors' activities by looking at their computer screens, which typically displays an Orphanet web page related to the disease or family of disorders the ongoing phone call seems to be about. Orphanet provides in this context one of the most reliable and comprehensive information sources for scientific, medical, healthcare and support dimensions of rare diseases. This allows MRIS counsellors to identify for example which hospital hosts the relevant expert centre for establishing or confirming diagnosis, which patient association can provide support, or to relay specific procedures for obtaining drug coverage. As one MRIS counsellor puts it:

> "Thank God we have Orphanet, it's our main source of information! When Internet access is down or there is a problem with the database, we feel very destitute. I can't remember all the details about thousands of diseases, let alone all the different information related to them! [...] We had an in-house database before, but we have stopped maintaining and using it now that we use Orphanet documents."

This important reliance on Orphanet for counselling activities often raises specific issues. First and foremost, interactions on the help line rarely take the form of pure "information delivery" about a well identified disease entity. In many cases, people calling are preoccupied by health problems for which they (or a close relation) still haven't obtained a medical diagnosis. Moreover, as one can imagine, these interactions are fraught with emotional aspects, which entail considerations on the phrasing and content of rare disease descriptions in Orphanet documents, and how to manage potentially alarming information. This is well illustrated by the following statement made by another MRIS counsellor:

> "Orphanet is highly ranked in search engine results. When people enter the name of a [rare] disease, chances are they'll come across a related page on Orphanet, and read it. But they still have questions and want to talk about it. So we have to deal with the reactions of people who read these documents… which can be useful, you know, to improve them… Because, there are at least some documents that in my opinion are problematic… the way things are said, it's harsh sometimes… Of course they need to be comprehensive, but sometimes it's not good to be too comprehensive. This is striking for example in some pathologies with prenatal diagnosis, which are very distressing situations. The person does not know her baby, no one has seen the baby, really… and the document describes a list of things that can be very distressing for people, among which some appear in only 1% of the cases."

Over the years, what was a strong collaboration between MRIS counsellors and Orphanet staff, taking place through weekly face-to-face meetings, declined to less frequent exchanges through email. MRIS participation to discussions concerning for instance which disease entities should come first in Orphanet's work priorities, based on their experience at the help line, slowly dwindled. In the space of a decade, most of the diseases that MRIS counsellors encountered in the course of their activity – typically the most frequent of rare diseases – had already been integrated to the database, and well documented, according to Orphanet's standards. One gets the impression that MRIS counsellors consider that over time, their growing reliance on Orphanet has come to the detriment of their active involvement with the organisation, and the taking into account of some of their specific concerns and forms of expertise.

Infrastructural Inversion as Strategy and Practice

Bowker (1994) has introduced the notion of "infrastructural inversion" as a methodological lens which allows the analyst to capture how things like databases and classification systems are embedded into the practices of collectives that share a common infrastructure. By bringing to the fore the mundane technical and organisational processes that sustain an infrastructure's operation, one can bring back to light important aspects that ordinarily tend to recede into taken-for-grantedness. Drawing on this insight in his masterful work on the climate knowledge infrastructure, Paul Edwards (2010: 20) argues that infrastructural inversion provides more than a methodological tool available to the ethnographer in the field. He finds that the scientists he studies also resort to this strategy as they negotiate infrastructural commitments, and that “infrastructural inversion is, in fact, fundamental to how scientists handle data”. We contend that this argument is not limited to scientists, but holds true for many actors engaged in knowledge work of some kind. This can be empirically illustrated through observations of “classification meetings” that take place on a monthly basis at the Rare Diseases Platform.

In order for a rare disease entity to appear inside the relational database, it must at least be given a place in a biomedical classification, and be linked to a set of resources (a written definition, at the very least). During classification meetings, the Orphanet director, the scientific director, some information scientists, as well as invited members of the Platform gather to go over the entities whose creation in the database should be given priority, as well as those which, in relational database parlance, need to be “deleted”, “dis-included” or “unlinked”. Collective discussions and decisions rely on documents compiled during the abovementioned research phase, and take into account general clinical and scientific considerations, but also attend to more practical emergencies, based especially on a list of requests coming from other partners in the Platform or from scientists working worldwide on genes involved in certain rare diseases.

These meetings provide an occasion to witness how political, moral and epistemic principles are collectively debated, problematised and articulated with technical considerations. For example, the heading under which a disease will appear in the database (alongside a list of synonyms that link back to the main entry) is deemed crucial. Some proposed names don’t make good candidates because they don’t “sound” right, or look too complicated, prompting reactions such as: “You always need to think about the patient […] A disease name is used when patients, doctors, and other people talk to each other, after all, not just fundamental researchers…” Other important matters are also addressed, such as the group in which a particular disease should be placed among the different available clinical and genetic classifications. Another topic pertains to the types of documents that should be produced: will just a summary do (but “even writing a summary involves touching everything in the database”), or is it a disease that “deserves a real text” and other resources to be added (links to scientific publications, lists of genes involved, etc.)? Should a particular disease be included in the database but left “mute”, meaning that it will be recorded but not accessible via the web portal since it is not linked to any resource? Occasionally, for a host of reasons, some diseases need to be “dis-included”: a single entity is split into two or more entries, each with its own resources. Finally, some diseases are deleted when, for example, they have remained unlabelled in the database for too long and no resources have been linked to them. Decisions in creating, including, deleting a disease, as well as managing the diversity of available resources – and their consequences for the database’s very development – appear very tricky, because they have meaningful impacts everyone needs to be aware of. In the course of these meetings, those involved seek to reaffirm the priorities behind Orphanet’s and the Platform’s missions, as illustrated by the following arguments: “It is the clinical aspect that counts in the long run”; “It makes no sense to create a disease if you do not have a good text to go along with it”; “We need to be careful about this deletion, it would make no sense to lose knowledge!”

These vignettes underline the fact that making multiple perspectives on rare diseases converge in meaningful, compatible and efficient ways is not an easy task. As negotiations unfold, priorities are established and reaffirmed, and some values gain precedence. This results in specific ways of framing rare diseases being better represented than others, both on the private side of the database, and its public manifestation through the web portal. The way navigation through the Orphanet portal is entirely framed and constrained by disease entities is a case in point. Moreover, given the prevailing perspectives of the actors of the Platform, clinical considerations remain prominent, at the expense, for example, of etiological considerations. In performing this kind of categorical work, actors rely on the relational database, and discussions about its inner workings and evolution, as a fulcrum to articulate work processes within the boundary infrastructure. The database therefore appears in its dual character as means and end: the focus of people's work, it is the reason for which they work (to enhance and maintain the database); but as a work tool, it is also the instrument with which they strive to fit together scales, temporalities and outputs within the infrastructure.

Here then, we witness infrastructural inversion taking place for two main reasons. The first one has to do with the fact that gaining recognition and securing resources for database-related activities poses a genuine problem in terms of representing the work this involves and its outcomes. The various tasks required to collect, connect, edit or delete data - and the motives underpinning these activities - are often poorly understood, under-valued or even ignored, and end up being relegated to "mere" maintenance and updating activities. Infrastructural inversion, then, is a means to restore the complexity of these activities, and reaffirm their crucial role, by unfolding the heretofore invisible technical and organisational intricacies that sustain them (Star & Strauss, 1999; Dagiral & Peerbaye, 2012). The second reason is related to the fact that most actors we met were well aware of the provisional character of knowledge work, and recognised that the procedures one chooses to validate knowledge, the types of information one favours, the ways in which data are linked and presented all play a role in the kinds of knowledge that are given pre-eminence, which in turn shape forms of commitment and collective action. Infrastructural inversion, then, is also a strategy that partakes in the negotiation of hierarchies and priorities, generates controversies, reaffirms convictions, and carries weight in decisions. In this respect, we see infrastructural inversion happening not only when things break down and need to be repaired (Star & Ruhleder, 1996) but also when the infrastructure needs to be updated, upgraded or extended, all activities that require some form of vigilance, maintenance and care (Denis & Pontille, 2015).

Beyond the Database: Caring for the Boundary Infrastructure

Keeping It Infrastructural: Maintenance and Extension

The allocation of repair and maintenance work across the Platform brings to light a number of elements pertaining to what needs to be done in order for a boundary infrastructure to function satisfactorily for its constituency. This might be related for instance to the upgrading of electrical installations (*e.g.* setting up uninterruptible power supply units for safeguarding computer workstations) or computer network administration tasks – a consideration which highlights the need for infrastructural commitments to articulate and bridge a varied set of technological requirements (Vertesi, 2014). As one of the organisations staffed with the most "computer specialists", Orphanet was particularly solicited in this respect, much to the dismay of the concerned individuals, who regularly complain about how poorly boundaries between the work of programmers, web developers, database managers, and computer systems administrators are understood and respected across the whole platform. An additional illustration is provided by the many tribulations of Orphanet's hardware servers, housing working versions of the relational database. The story, as recounted by one of the database managers, exemplifies challenges related to the work of building conventions and of establishing stable relations with things, especially in a context of organisational and institutional fragility (high staff turnover, frequent relocations and so on):

> "[Until the early 2000s] our machines were hosted at Infobiogen [*a now defunct French bioinformatics resource centre*]. Then Infobiogen closed down, and our machines wandered about a bit for some time. We found a temporary solution as 'guests' of the Paris IV University, but the accommodation was not on par with what we had before, and we needed more resources. Then we had the opportunity of being hosted within Inserm's DSI [*the general IT Systems Department*], which means we integrated a more 'normative' environment. And this is far from being insignificant, from a cultural perspective. Because at Orphanet, from the very beginning, we have had a very tinkering, hands-on approach to things – in the best sense. [...] When you arrive in an

> environment like the DSI, things are different, there are dedicated teams, infomanagement services and so on… You can't have remote access to your server in exactly the same way. You need to log in with credentials, fill forms explaining what, why… It's very tedious and time-consuming. But at the same time, when you don't do things this way, you can run into problems… Like getting hacked by a Turkish pirate [*laughs*] because the update was not done properly, because the engineer… well the engineer was leaving [Orphanet] and had not done things correctly, and the architecture was not strong enough. These things happen. And how you need to act in these situations, is by implementing a number of processes that will entrench action independently of individuals. Because the problem is, individuals come and go, and it is dangerous to have a system whose processes depend entirely on individuals. It's great having people, you need motivated people, that's for sure, but you can't rely only on them, because you can put yourself in danger for trivial reasons."

The division of labour across organisations points out the tensions related to forms of taken-for-grantedness or invisibility, and the distributed ways of managing and taking care of shared spaces, tools, technologies and perspectives, further emphasising the intensely relational nature of infrastructure. However this dimension of an infrastructure's daily life refers to only one aspect of what caring about any boundary infrastructure is actually about. Building upon work dedicated to the study of large technical systems, and especially cities, Steven Jackson (2014: 231) suggests that:

> "foregrounding maintenance and repair as an aspect of technological work invites not only new functional but also *moral* relations to the world of technology. It references what is in fact a very old but routinely forgotten relationship of human things in the world: namely, an ethics of mutual care and responsibility."

If maintenance and repair usually sends us back to the present and its urgency, here above all else the care of things reflects the necessity of anticipation, of thinking about future developments and what can be gathered under the term of "extension", referring to the work of infrastructuring foresight and extensibility. Maintenance work as "care" therefore does not only equate with maintaining the installed base in a good state of upkeep or repair: what is at stake here has to do with ensuring continuity and reconfiguration, robustness and flexibility, as technologies and commitments evolve, and the transformations of initial projects call forth new versions of the system, and generate new meanings for collective endeavour. One good example of this lies in Orphanet's efforts, dating from around 2005 onwards, to initiate developments in order to interface its database with other online reference resources in the biomedical world, such as SwissProt, OMIM, MedDRA, Snomed, MeSH and the catalogue and index of French-language medical sites (Cismef) maintained by the Rouen University Hospital Centre. In order to exist with ever greater relevance in this landscape of health information and biomedical research, and to ensure the visibility of rare diseases therein, Orphanet's development strategy started giving crucial importance to this work of extension, interconnection and standardisation. Maintenance work therefore tended towards the anticipation and construction of extensions rendered necessary by the shifts within the boundary infrastructure. In order to address concerns related to connectability and extension in the project's orientations, which remain largely unpredictable due to uncertainties regarding future technological trends, the technical team needs to embed this work in the standardisation processes at play in the IT worlds. From this point of view, the tension between standardisation and flexibility which is characteristic of many knowledge infrastructures becomes visible (Hanseth et al., 1996). This is not so much a case of interpretive flexibility as what could be termed flexibility "by design", intended to fulfil the promise of securing data and their relations irrespective of the paths subsequently taken.

Negotiating New Boundaries

The observation of meetings devoted to a project on "multi-technology health services" provides an opportunity to grasp the fragile balance between "legacies" of past versions of the database and the promise of "shareable ontologies" afforded by the semantic web, pointing to potentially new ways of interconnecting with external databases as well as generating classifications. The development of projects such as these requires interfacing entire parts of the database with other bases, each with its own history, made up of technical choices, norms, standards and particular specifications, which must be taken into account for effectiveness. At the same time, the normalisation and standardisation work inherent to database management systems as a field is largely geared towards such objectives of integration, fusion and interoperability. A consultant specialised in the domain of biomedical databases has been advising the Orphanet team on this matter since the organisation first purchased a Sybase licence in 1996. Gathered around whiteboards displaying hand-drawn graphic representations of the database, the consultant and four of the members of the IT team regularly remind newcomers in their discussion of why a given element of the database is like that "for historical reasons", or justify the nature of the relations between elements inside a data table.

In order to ensure its financial stability, and renew its relevance as a member of the Platform, Orphanet has also gradually been led to build a number of commercial and public services designed to promote its data, supplying specialised extractions of its database for the worlds of clinical and fundamental research as well as the pharmaceutical industry. This has entailed establishing new working priorities in order to focus on disease entities that were of topmost relevance to biomedical institutions and drug companies (namely diseases for which ongoing clinical trials, drug development or research projects were available).

Furthermore, as from 2010, and as a consequence of the Platform's involvement in the European Commission Expert Group on Rare Diseases, the disease classifications produced within Orphanet gained recognition as a central reference for the inclusion of rare diseases into the next revision of the International Classification of Diseases (ICD-11) maintained by the World Health Organisation (WHO). Bowker and Star (1999: 107-108) had already noted in their analysis of the ICD as a constantly evolving boundary infrastructure how much this central tool depends on the interaction between diverse preexisting infrastructures (databases, classification systems, national and international institutions). At the level of the Platform, new commitments and reconfigurations had therefore to be negotiated, in order to ensure that the database would be able in the long run to supply the WHO with classifications and ontologies in a novel form, with a view to integrating rare diseases into the ICD. This task required that members of the boundary infrastructure devote time and energy to reshape the classifications and annotations embedded in the relational database, in order to ensure their compliance with the ICD requirements. One immediate consequence of this was a slowing down of the general rate at which the various electronic resources associated to disease entities were being added or updated in the relational database.

These examples highlight the dynamics of belonging, inclusion (and sometimes exclusion) that go along with infrastructures as "invested forms". They also suggest that oftentimes investment in form is not a process that happens at the outset, as a prerequisite that will be followed by periods of relatively unproblematic coordinated action. Rather, as the boundaries of the infrastructure shift – being both a cause for and a consequence of investment in form – this in turn calls for further adjustments: investment in form builds on and reconfigures previous investments. This is made manifest through the intertwined regimes of care that are part and parcel of maintaining and reconfiguring a boundary infrastructure like the one represented by the Rare Diseases Platform.

Conclusion

Exploring the life and work of the knowledge infrastructure that constitutes and is constituted by the European rare diseases community, one cannot help but being struck by its dual nature, as something which alternates between being "taken for granted" and "problematic", "learned as part of membership" and open to negotiation. Both from an external perspective, and from the insider outlook of the actors of the Rare Diseases Platform, infrastructure embodies both an achievement that manages to *represent* rare diseases and its multiple challenges, and a project which in its very form is fraught with enormous fragility and uncertainty – which in turn weighs upon the work of those involved and never ceases to question their collective involvement. This constant shift between taken-for-grantedness and problematicity, between learning and calling into question, is probably a more general feature of knowledge infrastructures, especially when they exist as a modality of boundary infrastructures encompassing multiple causes. In particular, the fragility of commitments as well as the work required to maintain the meanings of collective action play an important role in this essential tension. In this respect, the impossibility to reach a state of completion, given the levels of flexibility and open-endedness required, is perhaps a key component of a knowledge-in-the-making-oriented infrastructure.

Our ethnographic study of the Rare Diseases Platform therefore provides insights that further our understanding of knowledge production issues in projects that are not limited to scientific outputs, but also include explicit political and moral objectives. In this paper we have shown how what counts as knowledge is negotiated between the participants of a boundary infrastructure, which relies centrally on a relational database situated at the crossroads of a large and complex network of individuals, institutions and practices. Epistemic, political and ethical commitments are materially enacted within the sociotechnical framework provided by the database through an ecology of work practices that shape ways of knowing, living with, and fighting social ignorance of, rare diseases.

We have also analysed infrastructural inversion as a practice that members of the boundary infrastructure resort to in order to negotiate the meanings of their collective action. It allows them to articulate knowledge-related activities with other forms of mobilisation, that do not necessarily rely on the database itself, and their active involvement in its inner workings. Infrastructural inversion allows the tension between local implication at the level of the database and global engagement for rare diseases as a political cause to be reflexively and productively managed. We contend that the existence of the Rare Diseases Platform as a boundary infrastructure ultimately rests on this process.

Furthermore, in the empirical case under scrutiny, what is at stake in the building and maintaining of a boundary infrastructure devoted to the cause of rare diseases extends well beyond issues of coordination or scientific collaboration. As Keating and Cambrosio (2003: 324) argue: “insofar as they embody regulations and conventions of equivalence, exchange, and circulation [...], platforms are not simply one among many forms of coordination that include networks; rather, they account for the generation of networks or, at the very least, they are a condition of possibility for the very existence and transformation of networks”. The Rare Diseases Platform as a boundary infrastructure manifests itself as a specific network for representing rare diseases, that reflects situated material cultures and political agendas, and is generated through particular investments in form. These investments evolve in time, as technologies change, and different projects are carried out (Karasti et al., 2010). Consequently, infrastructural commitments need to take account of past legacies and anticipate new requirements and visions. The shape of the boundary infrastructure as a whole is thereby in a state of continued reconfiguration.

The perspective adopted in this article tends to play down the extended involvement of Orphanet and members of the Platform in other spaces where research on and management of rare diseases take place, such as the European Union Committee of Experts on Rare Diseases, the European Medicines Agency, or the International Rare Diseases Research Consortium. However, we suggest here that what happens in these arenas must be explored and understood in the light of the knowledge work being produced within the boundary infrastructure. How curators tackle sociotechnical issues inside and beyond the database ultimately shapes the broader agenda of rare disease policy.

Notes

1 http://www.orpha.net/

Acknowledgements

This research was in part supported by the Ile-de-France Regional Council (Institution-Citizen Partnerships for Research and Innovation grant, PICRI 2008-2011). We warmly thank the special issue editors and three anonymous referees, whose comments and suggestions were most helpful for improving this paper. We are especially grateful to Madeleine Akrich, Florence Millerand, Jérôme Denis and David Pontille for providing opportunities and food for thought to keep conversations about infrastructures alive.

References

Aymé S & Rodwell C (eds) (2013) 2013 Report on the State of the Art of Rare Disease Activities in Europe. European Union Committee of Experts on Rare Diseases. July 2013.

Bietz MJ, Baumer EPS & Lee CP (2010) Synergizing in Cyberinfrastructure Development. *Computer Supported Cooperative Work* 19: 245-281.

Bowker GC (1994) *Science on the Run: Information Management and Industrial Geophysics at Schlumberger, 1920-1940*. Cambridge, MA: MIT Press.

Bowker GC (2000) Biodiversity Datadiversity. *Social Studies of Science* 30(5): 643-683.

Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Bowker GC, Baker KS, Millerand F & Ribes D (2010) Towards Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: Hunsinger J, Klastrup L & Allen M (eds) *International Handbook of Internet Research*. Dordrecht: Springer, 97-117.

Bowker GC & Star SL (1999) *Sorting Things Out: Classification and Its Consequences*. Cambridge, MA: MIT Press.

Dagiral E & Peerbaye A (2012) Les mains dans les bases de données : connaître et faire reconnaître le travail invisible. *Revue d'anthropologie des connaissances* 6(1): 191-216.

Dagiral E & Peerbaye A (2013) Voir pour savoir. Concevoir et partager des "vues" à travers une base de données biomédicales. *Réseaux* 178-179: 163-196.

Denis J & Pontille D (2015) Material Ordering and the Care of Things. *Science, Technology & Human Values* 40(3): 338-367.

Dutton WH & Jeffreys PW (eds) (2010) *World Wide Research. Reshaping the Sciences and Humanities*. Cambridge, MA: MIT Press.

Edwards PN (2010) *A Vast Machine. Computer Models, Climate Data, and the Politics of Global Warming*. Cambridge, MA: MIT Press.

Gieryn T (1983) Boundary-Work and the Demarcation of Science From Non-Science: Strains and Interests in Professional Ideologies of Scientists. *American Sociological Review* 48(6): 781-795.

Hanseth O, Monteiro E & Hatling M (1996) Developing information infrastructure: The tension between standardization and flexibility. *Science, Technology & Human Values* 21(4): 407-426.

Hilgartner S (1995) Biomolecular Databases: New Communication Regimes for Biology? *Science Communication* 17(2): 240-263.

Hine C (2006a) Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. *Social Studies of Science* 36(2): 269-298.

Hine C (ed) (2006b) *New Infrastructures for Knowledge Production: Understanding EScience*. Hershey: Information Science Publishing.

Huyard C (2009) How did uncommon disorders become "rare diseases"? History of a boundary object. *Sociology of Health and Illness* 31(4): 463-477.

Huyard C (2012) *Rare. Sur la cause politique des maladies peu fréquentes*. Paris: Presses de l'EHESS.

Jackson SJ (2014) Rethinking Repair. In: Gillespie T, Boczkowski PJ & Foot KA (eds) *Media Technologies. Essays on Communication, Materiality, and Society*. Cambridge, MA: MIT Press, 221-307.

Knorr Cetina K (1999) *Epistemic Cultures. How the Sciences Make Knowledge*. Cambridge, MA: Harvard University Press.

Karasti H, Baker KS & Millerand F (2010) Infrastructure Time: Long-term Matters in Collaborative Development. *Computer Supported Cooperative Work* 19: 377-415.

Keating P & Cambrosio A (2003) *Biomedical Platforms. Realigning the Normal and the Pathological in Late-Twentieth-Century Medicine*. Cambridge, MA: MIT Press.

Law J (1987) Technology and Heterogeneous Engineering: The Case of Portuguese Expansion. In: Bijker W, Hughes T & Pinch T (eds) *The Social Construction of Technological Systems*. Cambridge, MA: MIT Press, 111–134.

Lee CP (2007) Boundary Negotiating Artifacts: Unbinding the Routine of Boundary Objects and Embracing Chaos in Collaborative Work. *Computer Supported Cooperative Work* 16: 307–339.

Leonelli S (2013) Integrating Data to Acquire New Knowledge: Three Modes of Integration in Plant Science. *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 44(4): 503-514.

Leonelli S (2015) What Counts as Scientific Data? A Relational Framework. *Philosophy of Science* 82(5): 810-821.

Miller CA & Edwards PN (eds) (2001) *Changing the Atmosphere. Expert Knowledge and Environmental Governance*. Cambridge, MA: MIT Press.

Millerand F (2011) Le partage des données scientifiques à l'ère de l'*e-science* : l'instrumentation des pratiques au sein d'un collectif multidisciplinaire. *Terrains & Travaux* 18: 215-237.

Millerand F & Bowker GC (2009) Metadata Standards. Trajectories and Enactment in the Life of an Ontology. In: Lampland M & Star SL (eds) *Standards and Their Stories. How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Ithaca and London: Cornell University Press, 149-176.

Olson GM, Zimmerman A & Bos N (eds) (2008) *Scientific Collaboration on the Internet*. Cambridge, MA: MIT Press.

Parker JN, Vermeulen N & Penders B (eds) (2010) *Collaboration in the New Life Sciences*. Farnham & Burlington: Ashgate Publishing Company.

Star SL (2010) This is Not a Boundary Object: Reflections on the Origin of a Concept. *Science, Technology & Human Values* 35(5): 601-617.

Star SL & Griesemer JR (1989) Institutional Ecology, "Translations" and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science* 19(3): 387-420.

Star SL & Ruhleder K (1996) Steps Towards an Ecology of Infrastructure: Design and Access for Large Information Spaces. *Information Systems Research* 7(1): 111-134.

Star SL & Strauss A (1999) Layers of Silence, Arenas of Voice: The Ecology of Visible and Invisible Work. *Computer Supported Cooperative Work* 8: 9-30.

Strasser B (2011) The Experimenter's Museum: GenBank, natural history, and the moral economies of biomedicine. *Isis* 102(1): 60-96.

Thévenot L (1984) Rules and Implements: Investment in Forms. *Social Science Information* 23(1): 1-45.

Thévenot L (1986) Les investissements de forme. In Thévenot L (ed) *Conventions* économiques. Paris: Presses Universitaires de France, 21-71.

Thévenot L (2009) Postscript to the special issue Governing Life By Standards: A View from Engagements. *Social Studies of Science* 39(5): 793-813.

Vertesi J (2014) Seamful Spaces: Heterogeneous Infrastructures in Interaction. *Science, Technology & Human Values* 39(2): 264-284.

Studying the History of Social Science Data Archives as Knowledge Infrastructure

Kalpana Shankar

School of Information and Communication Studies, University College Dublin, Ireland / Kalpana.shankar@ucd.ie

Kristin R. Eschenfelder

School of Library and Information Studies, University of Wisconsin-Madison, USA

Greg Downey

School of Journalism and Mass Communication, University of Wisconsin-Madison, USA
School of Library and Information Studies, University of Wisconsin-Madison, USA

Abstract

We map out a new arena of analysis for knowledge and cyberinfrastructure scholars: Social Science Data Archives (SSDA). SSDA have influenced the international development of the social sciences, research methods, and data standards in the latter half of the twentieth century. They provide entry points to understand how fields organise themselves to be 'data intensive'. Longitudinal studies of SSDA can increase our understanding of the sustainability of knowledge infrastructure more generally. We argue for special attention to the following themes: the co-shaping of data use and users, the materiality of shifting revenue sources, field level relationships as an important component of infrastructure, and the implications of centralisation and federation of institutions and resources. We briefly describe our ongoing study of primarily quantitative social science data archives. We conclude by discussing how cross-institutional and longitudinal analyses can contribute to the scholarship of knowledge infrastructure.

Keywords: social sciences, data archives, institutional sustainability

Introduction

In this paper, we map out a new arena of analysis for knowledge and cyberinfrastructure scholars: Social Science Data Archives (SSDA). SSDA are global information infrastructures that have been influential in the international development of the social sciences, research methods, and data standards in the latter half of the twentieth century. SSDA provide entry points to understand how fields organise themselves to be 'data intensive,' to examine the mutual shaping between specific research disciplines and knowledge infrastructures, and to study the evolution of a field's valuation of different types of data across time and place. Longitudinal studies of SSDA can increase our understanding of the sustainability of knowledge infrastructure more generally.

We provide a brief history of the link between SSDA and the quantification of the social sciences and then outline the conceptual background of examining SSDA as part of infrastructure studies. We argue for special attention to the following themes: the co-shaping of data use and users, the materiality of shifting revenue sources, field level relationships as an important component of infrastructure, and the implications of centralisation and federation of institutions and resources. We briefly describe our ongoing study of primarily quantitative social science data archives. We conclude by discussing how cross-institutional and longitudinal analyses can contribute to the scholarship of knowledge infrastructure.

A Brief History of Social Science Data Archives and Archiving

We define SSDA at two levels of analysis. First, they are individual data archives with particular institutional missions and physical homes. SSDA are also professional organisations and consortia that promote shared values, standards, and goals via collaborative projects, conferences, and publications. Both levels of SSDA have curated and provided access to quantitative social science data for over fifty years, and both continue to exist in many nations alongside flashier examples of cyberinfrastructure and open data repositories. SSDA predate both computers and the Internet so their long history provides an opportunity to examine the people, tools, and organisations that constitute infrastructure, and how that infrastructure has adapted over time to change. Further, the existence of SSDA and related professional organizations across nations allow for comparisons of national science strategies in relation to knowledge infrastructure.

Historically, the social sciences have enjoyed quantitative data archives since the 1940s with antecedents dating from the 1920s (Green & Gutmann, 2007). Initially the SSDA movement was driven by household spending surveys, census and political opinion data. For example, in the US, data archives such as the Interuniversity Consortium for Political and Social Research (ICPSR) and the Roper Center for Public Opinion Research began as elections and survey research projects, acquiring machine-readable data in the 1940s. Hoping to usher in a new era of comparative and longitudinal secondary research, their sponsors opened them to the general public in the 1950s and 1960s (Eulau, 1989; House, 2004; Scheuch, 2003). In the UK, a committee began work in 1963 to promote sharing of output from government social science surveys, and the UK Data Archive (UKDA) was established in 1966 (UK Data Archive, 2007). The 1960s and 1970s saw the establishment of over a dozen services and professional associations to coordinate quantitative data collection, to promote data sharing, and to educate students and scholars about quantitative and machine processing analysis methods (White, 1977).

While early data archiving – especially in political science and economics –focused on quantitative data, the rise of post-structuralism and critical approaches in the 1970s led to increased popularity of qualitative data. Non-digital qualitative data (e.g., documents, photos) have long been curated, but the first explicitly digital qualitative data archives often developed in conjunction with existing quantitative data projects. For example, Radcliffe College's Murray Research Archive was founded in 1976 with an explicit multi-method collection development goal but was later merged with Harvard's Institute for Quantitative Social Science (Altman, 2009). The UKDA's 1994 Qualidata project is often cited as a pioneer in qualitative data archiving and has become part of the larger predominately quantitative archive (Corti, 2005, 2011; Smioski, 2011).

Despite the prominence and longevity of SSDA, their history has never been told in a critical, synthesising way. SSDA figure surprisingly little in larger histories of the social sciences or knowledge infrastructures. Instead, like many stories of technological progress in the digital age, a thin and instrumental tale of institutional winners has survived in the literature of information sciences and survey methodology. The standard narrative, written almost entirely from the perspective of advocates, leaves out much of the contingency and anxiety involved in trying to build and maintain the technology, funding, expertise, and client base. It tends to represent relationships among archives as entirely positive, overlooking periods of competition and discord. Importantly, it also skirts the thorny question of why scholars might or might not choose to use SSDA in the first place – a question that remains raw as new and often uncurated sources of 'open data' proliferate. A deeper, STS-inflected critique of this standard narrative can contribute to current conversations on the sustainability of knowledge infrastructures because problems experienced by SSDA in the past are parallel to challenges facing newer cyberinfrastructure projects today.

Conceptual Background

Edwards (2010) defined knowledge infrastructures as 'robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds'. Today we easily understand infrastructure as encompassing networked digital data repositories, but pre-computer arrangements that supported the collection and distribution of data via punch cards and tape drives are also infrastructure. Knowledge infrastructures must also be understood as systems with both social and cognitive elements (Starr, 1987).

The story of the development of SSDA fits into what Porter (1995: xii) called 'the history of quantitative objectivity' and the desire for some social science fields to increase their stature and legitimacy (See also Porter, 1986). SSDA activities from the 1940s to the 1970s were motivated in part by the desire of some fields to conduct statistically based comparative and longitudinal analysis to improve the quality of their research claims. Use of these methods arguably raised the prestige of fields in the university environment (Porter, 1995). At a national level, political pressures led funding agencies to emphasise their support for more 'empirical' research focused on 'relevant' or 'practical' outcomes – preferences that favored support for quantitative data (Massey, 2000). Moreover, quantitative methods more easily extend scientific communities of practice across nations (Porter, 1995). The social science data archives movement promised just such a technology of distance to the practitioners of comparative survey and census research in different nations (Rokkan, 1979). The dream of creating a global 'demoscope,' as one sociologist dubbed it in the 1940s, promised a way of 'sampling the facts' across nations and time (Dodd, 1946).

SSDA proponents have long envisioned their institutions to be in the business of promoting sharing and reuse of research data; however, the practicalities of the enclosure and disclosure of data have been contentious in SSDA since their inception. Many scholars have envisioned a technological utopia of data sharing (Willinsky, 2000). However, other scholars have pointed to the complications of creating sustainable 'knowledge commons' and have argued that some enclosure/restrictions on release are necessary for sustainability (Ostrom & Hess, 2006).

A final conceptual framework synthesises the earlier frames as set of all-encompassing 'practices' for organising memory. Bowker (2005) writes that a system of saving information, or 'memory practices' are both 'sequential' in that they declare a starting point for a new set of values and protocols in preserving the past, and 'jussive' in that they necessarily define and even value what can and can't be saved (Bowker, 2005: 228). Further, memory practices are always contested and evolving even as they claim to preserve unchanging understandings across time and space. Telling the history of SSDA reveals such contingency and jockeying, as new memory practices are developed and deployed in an environment of rapidly changing technologies and constantly shifting relationships.

SSDA and the Evolving 'Social Sciences'

Across different arenas, calls for rational and scientific decision making led to increased demand for personnel with the skills to work with quantitative data and its associated methodologies. As Igo (2007: 5) described, 'Professional statisticians, government bureaucrats, academic social scientists, and all manner of planners claimed that survey methods, newly 'scientific,' were essential' for understanding change, for 'managing a complex industrial society'.

The desire for increased rationality in decision-making also led to increased support for quantitative data archiving and education. As early as 1929, the Social Science Research Council (SSRC)[1] included among its key objectives data preservation, dissemination of 'materials, methods and results,' and 'reproducing basic data' to increase access and encourage reuse (Carpenter, 2007).

Early SSDA proponents argued for increasing institutional capacity to support scholars and decision makers in using social science methods and quantitative data (Miller, 1989: 152). Miller (1989: 152) envisioned a social science data service that would 'make a massive difference in the development of the research skills of individual scholars' and provide 'institutional support… like the equivalent of the traditional scholar's library and reference service'.

The push for quantified social science research was also related to US federal funding for research and the desire for political approval of research (Solovey, 2013). Solovey (2013: 4) describes the early lack of support for social science research in the National Science Foundation (NSF) during the 1950s as stemming from a 'scientism' which saw credibility and progress as tied to 'rigorous, systematic, and quantitative investigations'. During Congressional hearings on the founding of the NSF, witnesses from the SSRC 'all accepted a sharp distinction between objective social research on the one hand and social reform, political ideology, and value-laden inquiries on the other hand' (Solovey, 2013: 22). Sociologist Harry Alpert, who worked at the NSF from 1953 to 1958, 'crafted a carefully circumscribed strategy that limited NSF support to the so-called 'hard-core' end of the social research continuum' (Solovey, 2013: 149). It was during this time that Alpert (1956) linked such principles of objectivity to the very survey data that the nascent SSDA movement would soon rally around.

Two key questions around labour arose in the evolution of SSDA and their relationship to the social sciences (and these questions persist today). The first concern was related to the education and employment of social science researchers. Would their career prospects be better served if they engaged in the slow and costly gathering of primary data or focused on the more rapid and cost effective use of secondary data? Was choosing to specialize in secondary quantitative data analysis a valid career option for a new social scientist, or did prestige require the collecting of one's own data? A second concern: who should have power over and responsibility in and for data archives? Social science professors collect and use the data from archives and may set collection priorities, but the day to day labour of running archives is not rewarded in most scholarly fields. The field of 'data librarianship' embraced the responsibility for data archives and archiving work, but lacked direct connection to the scholarly fields represented in the data and access to resource-level decision makers. Finally, computer expertise was needed to set up systems and define solutions for digital preservation, but ongoing systems maintenance was often not rewarded as innovative work in computing research (Bisco, 1966; Nasatir, 1973).

Co-Shaping of Use and Users

The history of SSDA provides an excellent frame with which to examine the mutual shaping of users and technologies through consumption, modification, disuse or reconfiguration of infrastructure (Oudshoorn & Pinch, 2005). Lack of use (either the use of data or the deposit of data) remains a prominent concern for infrastructure (Meyer & Schroder, 2015) and SSDA have a long history of trying to create and maintain user bases and supporters through adjustment to shifting fashions in social science. Further, in order to encourage development of the social science and their base of users and supporters, most SSDA missions include education to use quantitative methods that rely on archived data. For example, while SSDA were often initially driven by narrower missions to collect and preserve electoral, census and public opinion data, numerous examples exist of how SSDA adjusted collection development goals in response to changing scholarly demand for different types of data (e.g., qualitative, economic, gender/race, health) and then provided outreach and training materials on the newly-acquired data sets. Also, numerous examples exist of SSDA approaching new audiences to develop new users (such as local government officials, government agencies, community colleges, and even secondary schools). Similar strategies have been articulated in other knowledge infrastructures. For example, Ribes and Bowker (2009) note that in the GEON project, the participants' learning and training about the informational dimensions of their science reoriented their stance towards their broader intellectual community.

Material Concerns: Paying the Bills

While information infrastructures are highly diverse, they share many material concerns. One ever-present concern is how to pay the bills. The SSDA community includes organisations with a variety of revenue models: subscription-based, contract revenue, core government funding, and others (Kitchin et al., 2015). It is important that scholars critically examine the business models and revenue streams of infrastructure projects and how they change over time (Eschenfelder, 2014). Studying changes in business models potentially helps explain concomitant changes in relationships (i.e. user communities, host institutions, funders, competing data services) and technology infrastructure.

A related opportunity for analysis is how to package data to attract funders, users, and depositors. Questions of revenue are directly tied to decisions about how to package data, what types of access or use restrictions to put in place, and what types of services to offer. Because SSDA operate under a variety of revenue models, some openly share data, while others have data sets that are restricted to paying members. Some offer unprocessed data openly, but require membership for access to cleaned data. This diversity suggests that there are a variety of models under which SSDA might design configurations of data and services that attract users, induce data owners to deposit high-quality data, and retain the support of member payments, funders or other financial supporters. Examination of the SSDA field's 40 plus years of design decisions about data products and services provides a highly contextualised analysis point through which to understand the complexities about data access that go beyond simplistic dualisms of completely open and unfortunately enclosed, and to consider the shifting technoscientific, sociotechnical, and institutional forces shaping information infrastructure design decisions more generally.

Lastly, we suggest that understanding business models would allow STS researchers to incorporate theory and empirical findings from cognate areas such as economics (Ostrom, 1990), ecology (Brand & Jax, 2007), organisational theory (Lawrence et al., 2009), and production research (Bhamra et al., 2011) can give potential insights into the creation and maintenance of knowledge infrastructures. The lifespan, or sustainability, of infrastructure is a growing concern as researchers, archive practitioners and funders seek to ensure that resources invested in developing information institutions will have benefits that endure beyond the initial funding period (Crow, 2013; Ember & Hanisch, 2013; National Academy of Sciences, 2014; Maron & Loy, 2011a, 2011b). To contribute to the building of more sustainable infrastructure, scholars have called for greater understanding of the long-term technological, work organisation, and institutional development challenges faced by developers (Dutton & Meyer, 2010; Ribes & Finholt, 2009). Projects have begun to map out shifting technoscientific, sociotechnical, and institutional demands that influence sustainability of science infrastructure (Ribes & Polk, 2014).

Inter-SSDA Relationships and Field Level Infrastructure

Working from core STS prescriptions to examine relationships and networks of interactions, we propose attention to both local and field level infrastructure. While individual archives build and maintain infrastructure to provide access and use of social science data, we argue that networks of SSDA organisations represent an additional and equally important institutional layer of field-level infrastructure that both support and constrain individual SSDA. STS scholars should examine the field-level groups, their relationship to individual organisations, and their role in developing and supporting shared values, common practices and assumptions within a field. For example, professional organisations may recruit and indoctrinate new labor into taken-for-granted values or practices. They also may serve as a platform for shared projects such as standards development, provide a forum to disseminate new ideas, or give individuals a mechanism to develop influence in the community through governance and leadership. Consortia of organisations may pool resources to accomplish larger goals like standards development, make arrangements to divide work, spread field level practices into new nations through financial or other support, and advocate to governments or other funders for support of the broad goals of the field (i.e. the importance of data archiving in general). Importantly, field level organisations also compete with each other for the limited financial and attention resources of their member organisations and individuals working in the field.

Internal histories of the SSDA field are often triumphalist, focusing on the success of collaborative projects such as the Data Documentation Initiative, or the more contemporary DataPASS project (Gutmann et al., 2009; Vardigan et al., 2008) The field literature provides only hints about the tempestuous nature of inter-archive relationships that have over time included competition for territory (e.g., disciplinary, geographic), disputed use of each others' data, competition for extramural funding and project awards, and negative feelings stemming from perceived dominance of some archives during certain periods. The field literature about international consortia focuses on motivating participation and describing ongoing projects. It does not address or explain periods of lack of activity or unsuccessful initiatives.

Field level organizations are part of infrastructure in that they develop knowledge and practices, perform advocacy, create bridges to related groups, and provide resources and legitimacy. STS scholarship can add insights to infrastructure studies by examining competition between the field level groups for resources and prestige, considering less-successful group projects or periods of inactivity, and critical analysis of who does not participate in these organisations. Examples from the SSDA field include IASSIST, formed in the mid 1970s to represent social science data archive professionals globally (Heim, 1980; Adams, 2006; IASSIST, 2015). IASSIST states a goal of bringing together information professionals with social science researchers and computing specialists (IASSIST), but it is unclear to what degree IASSIST has succeeded in being a bridging infrastructure between these groups.

The field also includes several longstanding consortia of archives. These include the International Federation on Data Archives (IFDO) that coordinates field level goals and practices for SSDA internationally. In Europe, the Council on European Social Science Data Archives (CESSDA) has provided a coordinating platform for European projects.

Historical perspective also aids analysis of networks of organisations as infrastructure. Because participants may be more willing to talk about past conflicts than present tensions, it is easier trace the rise and dissolution of relationships over larger periods of time and it facilitates consideration of how shared values, taken for granted practices, and assumptions develop and are perpetuated or challenged over time.

Nationalisation/Federation of Infrastructure:

One analysis point relevant to all information infrastructure studies is whether infrastructure develops in a centralised or federated manner. Centralised infrastructures may provide national-level (or even international) services across multiple organisations in a field. Federated infrastructures include independent instantiations of infrastructure that may coordinate efforts in different ways over time. National research policies and funding patterns can directly influence the degree of centralisation or federation in infrastructure development (Rajabifard et al., 2006), leading to different types of field level coordinating organisations and different types of cooperative and competitive relationships within the field. Is it more effective to centralise infrastructure for a sub-field of research (or even for an entire country), or is it better to have multiple distributed manifestations of infrastructure? Exploring how such decisions are made and under what circumstances can reveal how actors mobilise resources, develop policy, collaborate and compete, and stabilise their institutions.

To give an example, SSDA infrastructure developed differently in Europe and the United States, and in various subfields of the social sciences. In Europe, and in many other parts of the world, national data archives have been common; for example, European nations send only one

SSDA representative to organisations like CESSDA and IFDO. In the United States and Canada, there has never been governmental recognition of a national SSDA; currently, seven different US archives are members of IFDO. But at various times stakeholders advocated for a vision of the national approach in North America. Further, at various times different US SSDA jockeyed for implicit recognition in the field as the 'US national data archive.' US social science data archive history however has seen repeated calls for creation of, and dedicated federal support for, a recognised national data archive or archives (Social Science Research Council, 1999; Ember & Hanisch, 2013). But, the fact that European nations send one representative to CESSDA may hide inter-archive competition for status and resources within nations. For example, while the UKDA has been the official representative to CESSDA, other SSDA have long existed in the UK.

Comparing Social Science Data Archives

We have argued that comparative STS oriented historical studies of social science data archives can provide insight into contemporary knowledge infrastructure concerns. Our own work in this area focuses on the history of five SSDA from the US and Europe and their relationship to each other from inception through the development of web based data access. For each of our sites, the archives' collection was founded with, and has historically primarily been dominated by, quantitative data. Our analysis will examine how each responded to increased demands for archiving of more diverse data types including qualitative data. From each case, we have collected historical documents such as annual reports, strategic plans, budgets and meeting minutes. We have also interviewed current institutional staff and past stakeholders. Our case sites include three of the most long-lived SSDA: the Interuniversity Consortium for Political and Social Research (ICPSR) at the University of Michigan and the UK Data Archive (UKDA) at the University of Essex[2] and LIS (formerly known as the Luxembourg Income Study). Our remaining case sites are all in relationship to ICPSR or UKDA.

To obtain a field level perspective, we have analysed documentation of the professional associations and consortia of data archives including historical materials of the International Association of Social Science Information Services and Technology (IASSIST), the International Federation on Data Archives (IFDO), and the Council on European Social Science Data Archives (CESSDA). We have traced networks of collaboration or competition between SSDA through analysis of relationships depicted in professional journals. We have also compared conversations about data archiving practices and data archiving labor as a profession in literature representing the voices of social scientists and information professionals. There are opportunities for other historical studies to examine other long-standing archives including for example qualitative data archives and linguistic data archives.

Conclusion

For decades, the social sciences have organised themselves to support long-standing domain data repositories – institutions with ties to specific scholarly fields that take on the mission of collecting, organising, preserving, and disseminating data for the purpose of furthering scholarship. They have drawn on archived data to produce longitudinal findings that would otherwise be impossible, and in recent years, large data sets for the types of prediction that have been en vogue across the disciplines. SSDA are an understudied 'space of convergence', which Chow-White and García-Sancho (2012: 125) define as 'technologically mediated processes of communication. They are the space of flows of people, disciplinary expertise, finance, cultural values, institutional ethics, technology, information, data, and code'.

In this paper, we urge greater attention on the part of STS scholars to SSDA as knowledge infrastructures for several interconnected reasons. First, their reach and influence across geographical boundaries are well-documented forerunners of today's cyberinfrastructures; their strategies for maintaining themselves over time have both pragmatic and scholarly implications. Secondly, they provide historical exemplars of how fields organised themselves to be 'data intensive', a persistent call that of course echoes today. SSDA serve as exemplars of how different disciplines have responded to the pervasive call for 'data intensive research' (Levallois et al., 2013). Lastly, SSDA provide a way to study across cases, time, and place in knowledge infrastructures; the availability of historical documentation, reports, and other working documents affords us opportunities to delve deeply into the interplay of budgets, staffing, and other daily administrative and institutional, often invisible labour, with the work of knowledge production and dissemination.

Lastly, STS inquiry could expand standard narratives of SSDA histories by paying attention to what Jackson (2014) argues is an important but undertheorised of knowledge infrastructures and design: breakdown and repair. We propose that other scholars bring to the study of knowledge infrastructure, as Jackson (2014: 222) eloquently writes, 'a deep wonder and appreciation for activities by which stability (such as it is) maintained'. The concept of breakdown and repair cuts across the themes we have described – not just the breakdown and repair of particular technologies and artifacts, but of institutions, relationships, even, potentially, the repair of memory itself. Paying attention to stories of discontinuity and rupture pave the way for understanding the 'mangle of practice' (Pickering, 1995). While such analyses may be methodologically and empirically complex, attention to these dimensions of SSDA (and other knowledge infrastructures) – how difficulties and discontinuities are worked with, worked through, and accounted for – would provide needed symmetry to our understandings of how knowledge infrastructures are created, managed, and ended.

Notes

1 References to the Social Science Research Council throughout this paper are to the US-based nonprofit organisation founded in 1923 to advance social science research and not to the ESRC's predecessor in the UK, which went by the same name.

2 The official name of the center is now UKDS (UK Data Service). The UK Data Archive label is retained by the University of Essex to describe the physical facility on the campus. The UKDS is a distributed archival service that includes data and services from other universities, including the University of Manchester. Since our archival research and interviews to date have focused exclusively on the Essex service and our study formally ends at 2002, we are retaining the UKDA terminology for the purposes of this paper.

Acknowledgements

The authors would like to thank the Alfred P. Sloan Foundation, the Irish Research Council, the Wisconsin Alumni Research Foundation, and ASIS&T History and Foundations SIG History Fund for financial support.

References

Adams MO (2006) The Origins and Early Years of IASSIST. *IASSIST Quarterly* 30(3): 5-15.

Alpert H (1956) Public Opinion Research as Science. *Public Opinion Quarterly* 20(3): 493-500.

Altman M (2009) Transformative Effects of NDIIPP: The Case of the Henry A. Murray Archive. *Library Trends* 57(3): 338-351.

Bhamra R, Dani S, & Burnard K (2011) Resilience: the Concept, a Literature Review and Future Directions. *International Journal of Production Research* 49(18): 5375-5393.

Bisco RL (1996) Social Science Data Archives: A Review of Developments. *American Political Science Review* 60(1): 93-109.

Bowker GC (2005) *Memory Practices in the Sciences*. Cambridge, MA: MIT Press.

Brand FS & Jax K (2007) Focusing the Meaning(s) of Resilience: Resilience as a Descriptive Concept and a Boundary Object. *Ecology and Society* 12(1).

Carpenter K (2007) Toward a New Cultural Design: The American Council of Learned Societies, the Social Science Research Council, and Libraries in the 1930s. In: Augst T & Carpenter K (eds) *Institutions of Reading: The Social Life of Libraries in the United States.* Amherst, MA: University of Massachusetts Press, 164-282.

Chow-White PA & García-Sancho M (2012) Bidirectional Shaping and Spaces of Convergence Interactions between Biology and Computing from the First DNA Sequencers to Global Genome Databases. *Science, Technology & Human Values* 37(1): 124-164.

Corti L (2005) Qualitative Archiving and Data Sharing: Extending the Reach and Impact of Qualitative Data. *IASSIST Quarterly* 29(3). Available at: http://www.iassistdata.org/downloads/iqvol293corti.pdf (accessed 24.1.2016).

Corti L (2011) The European Landscape of Qualitative Social Research Archives: Methodological and Practical Issues. *Forum Qualitative Social Research: Sozialforschung* 12(3): Art 11. Available at: http://nbn-resolving.de/urn:nbn:de:0114-fqs1103117 (accessed 30.1.2016).

Crow R (2013) *Sustainability of Open Access Services Report. Phase 3: The Collective Provision of Open Access Resources*. Available at: http://www.knowledge-exchange.info/Admin/Public/DWSDownload.aspx?File=%2fFiles%2fFiler%2fdownloads%2fOpen+Access%2fSustainabilityServices%2fSustainabilility+of+OA+Services+phase+3.pdf (accessed 30.1.2016).

Dodd SC (1946) Toward World Surveying. *Public Opinion Quarterly* 10(4): 470-483.

Dutton WH & Meyer ET (2010) Enabling or Mediating the Social Sciences? The Opportunities and Risks of Bottom Up Innovation. In: Dutton WH & Jeffreys P (eds) *World Wide Research: Reshaping the Sciences and Humanities*. Cambridge, MA: The MIT Press, 165-184.

Edwards PN (2010) *A Vast Machine: Computer Models, Climate Data, and the Politics of Global Warming.* Cambridge, MA: The MIT Press.

Ember C & Hanisch R (2013). *Sustaining Domain Repositories for Digital Data: A White Paper*. Ann Arbor, Michigan. Available at: http://datacommunity. icpsr. umich. edu/sites/default/files/WhitePaper_ICPSR_SDRDD_121113.Pdf (accessed 30.1.2016).

Eschenfelder KR (2014) Toward the Inclusion of Pricing Models in Sociotechnical Analyses: The SAE International Technological Protection Measure. In: Kleinman DL & Moore K (eds) *Routledge Handbook of Science, Technology, and Society*. New York: Routledge, 157-172.

Eulau H (ed) (1989) *Crossroads of Social Science: The ICPSR 25th Anniversary Volume*. New York: Agathon Press.

Green AG & Gutmann MP (2007) Building Partnerships among Social Science Researchers, Institution-Based Repositories and Domain Specific Data Archives. *OCLC Systems & Services* 23(1): 35-53.

Gutmann M, Abrahamson M, Adams M, Altman, M, Arms, C, Bollen K, Carlson M, Crabtree J Donakowski D, King, G, Lyle J, Maynard M, Pienta A, Rockwell R, Timms-Ferrara L, & Young C (2009) From Preserving the Past to Preserving the Future: The Data-PASS Project and the Challenges of Preserving Digital Social Science Data. *Library Trends* 57(3): 315-337.

Heim (McCook) K (1980) *Social Science Data Archives: A User Study.* Unpublished doctoral dissertation. University of Wisconsin, Madison.

House JS (ed) (2004) *A Telescope on Society: Survey Research and Social Science at the University of Michigan and Beyond*. Ann Arbor, MI: University of Michigan Press.

IASSIST (2015). *About IASSIST*. Available at: http://iassistdata.org/about/index.html (accessed 30.1.2016).

Igo SE (2007) *The Averaged American: Surveys, Citizens, and the Making of a Mass Public.* Cambridge, MA: Harvard University Press.

Jackson SJ (2014) Rethinking Repair. In: Gillespie T, Boczkowski PJ, & Foot KA (eds) *Media Technologies: Essays on Communication, Materiality, and Society*. Cambridge, MA: The MIT Press, 221-240.

Kitchin R, Collins S, & Frost D (2015) *Funding Models for Open Access Repositories.* Maynooth: Maynooth University and Dublin: the Royal Irish Academy and Trinity College Dublin.

Lawrence TB, Suddaby R, & Leca B (eds) (2009) *Institutional Work: Actors and Agency in Institutional Studies of Organisation*. Cambridge, UK: Cambridge University Press.

Levallois C, Steinmetz S, & Wouters P (2013) Sloppy Data Floods or Precise Social Science Methodologies? Dilemmas in the Transition to Data-Intensive Research in Sociology and Economics. In: Wouters P, Beaulieu A, Scharnhorst A, & Wyatt S (eds) *Virtual Knowledge: Experimenting in the Humanities and the Social Sciences.* Cambridge, MA: MIT Press, 151-182.

Massey D (2000) Editorial: Practising Political Relevance. *Transactions of the Institute of British Geographers*: 131-133.

Maron NL & Loy M (2011a) *Funding for Sustainability: How Funders' Practices Influence the Future of Digital Resources.* Available at: http://www.webarchive.org.uk/wayback/archive/20140614201525/http://www.jisc.ac.uk/publications/programmerelated/2011/fundingforsustainability.aspx (accessed 30.1.2016).

Maron NL & Loy M (2011b) *Revenue, Recession, Reliance: Revisiting the SCA/Ithaka S+R Case Studies in Sustainability*. Available at: http://sr.ithaka.org/sites/default/files/reports/4.4.2.pdf (accessed 30.1.2016).

Meyer ET & Schroder R (2015) *Knowledge Machines: Digital transformations of the Science and Humanities*. Cambridge, MA: The MIT Press.

Miller WE (1989) Research Life as a Collection of Intersecting Probability Distributions. In: Eulau H (ed) *Crossroads of Social Science: The ICPSR 25th Anniversary Volume*. New York: Agathon Press, 147-160.

Nasatir D (1973) *Data Archives for the Social Sciences: Purposes, Operations and Problems.* Paris: UNESCO.

National Academy of Sciences (2014) *Strategies for Economic Sustainability of Publicly Funded Data Repositories: Asking the Right Questions.* Available at: http://sites.nationalacademies.org/pga/brdi/pga_087151 (accessed 30.1.2016).

Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action.* Cambridge, UK: Cambridge University Press.

Ostrom E & Hess C (2006) A framework for analyzing the knowledge commons. In: Hess C & Ostrom E (eds) *Understanding Knowledge as Commons: From Theory to Practice*. Cambridge MA: The MIT Press, 41-81.

Oudshoorn N & Pinch T (eds) (2005) *How Users Matter: The Co-Construction of Users and Technology*. Cambridge, MA: The MIT Press.

Pickering A (1995) *The Mangle of Practice: Time, Science, and Agency*. Chicago: University of Chicago Press.

Porter TM (1986) *The Rise of Statistical Thinking, 1820-1900*. Princeton, NJ: Princeton University Press.

Porter TM (1995) *Trust in Numbers: The Pursuit of Objectivity in Science and Public Life*. Princeton, NJ: Princeton University Press.

Rajabifard A, Binns A, Masser I, & Williamson I (2006) The Role of Sub-National Government and the Private Sector in Future Spatial Data Infrastructures. *International Journal of Geographical Information Science* 20(7): 727-741.

Ribes D & Polk JB (2014) Flexibility Relative to What ? Change to Research Infrastructure. *Journal of the Association for Information Systems* 15(1): 287-305.

Ribes D & Bowker GC (2009) Between Meaning and Machine: Learning to Represent the Knowledge of Communities. *Information and Organisation* 19(4): 199-217.

Ribes D & Finholt TA (2009) The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Informaton Systems* 10(5): 375-398.

Rokkan S (ed) (1979) *A Quarter Century of International Social Science: Papers and Reports on Developments, 1952-1977*. New Delhi: Concept.

Social Science Research Council (1999) *Data Infrastructure for the Social and Behavioral Sciences: A Report to the Russell Sage Foundation on the Conference 'Wither Social Science Data and Data Infrastructure' Held at New York University, Nov 11-12*. New York University.

Scheuch EK (2003) History and Visions in the Development of Data Services for the Social Sciences. *International Social Science Journal* 55(177): 385-399.

Smioski A (2011) Archiving Qualitative Data: Infrastructure, Acquisition, Documentation, Distribution. Experiences from WISDOM, the Austrian Data Archive. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* 12(3): Art. 18, Available at: http://nbn-resolving.de/urn:nbn:de:0114-fqs1103181 (accessed 30.1.2016).

Solovey M (2013) *Shaky Foundations: The Politics-Patronage-Social Science Nexus in Cold War America*. New Brunswick, NJ: Rutgers University Press.

Starr P (1987) The Sociology of Official Statistics. In: Alonso W & Starr P (eds) *The Politics of Numbers.* New York: Russell Sage Foundation, 7-58.

UK Data Archive (2007) *Across the Decades: Forty Years of Data Archiving.* Essex UK: UK Data Archive. Available at: http://data-archive.ac.uk/about/archive/decades (accessed 30.1.2016).

Vardigan M, Heus P, & Thomas W (2008) Data Documentation Initiative: Toward a Standard for the Social Sciences. *The International Journal of Digital Curation* 3(1): 107-113.

White HD (ed) (1977) *Reader in Machine-Readable Social Data.* Englewood, CO: Information Handling Services.

Willinsky J (2000) *If Only We Knew: Increasing the Public Value of Social Science Research*. New York: Routledge.

Joe Deville, Michael Guggenheim, & Zuzana Hrdličková (eds) (forthcoming, 2016) Practicing Comparison: Logics, Relations, Collaborations. Manchester: Mattering Press. 199 pages. ISBN 978-0-9931449-0-5.

Steven D. Brown
sb343@leicester.ac.uk

It seems almost obligatory to start with a comparison. On the one side, the great sociologist Niklas Luhmann pulling cards from his filing cabinet, constructing texts which he claims are practically authored by the comparative process itself. On the other, the renowned anthropologist Eduardo Viveiros de Castro (2004: 14) provocatively asserting that 'It is only worth comparing the incommensurable, comparing the commensurable is a task for accountants, not anthropologists' (see Stockelova's chapter for more on this). It is the space in-between these two vivid images that this collection seeks to occupy.

Comparison, or perhaps simply the thought of comparison, seems to do strange things to those who engage in it (which as many of the contributors argue here, is practically all working social scientists at some point or another). If Luhmann can discern the logic of any given aspect of the modern world through rummaging in his card sorter, and Viveiros de Castro can argue for comparing practically anything with anything (just so long as the comparison is not initially obvious), then it seems that a strong orientation to comparison is a matter of academic reputation. Knowing what kind of 'comparator' one is, or the kinds of comparative practices one wittingly or unwittingly participates in, is described in this volume as a critical matter. In fact, as Krause argues, this is a more important matter than aligning oneself with any particular body of theory, since to think comparison properly is to think it outside of the constraints of theory.

A fun thought experiment to play with many edited collections is to imagine the authors sat together around a seminar table. Who will be the naughty one, deliberately provoking the others? Who will intervene in a calm measured tone, urging the need for synthesis? Who will sit there quietly furious, determined to have done with the conversation? Fortunately this is unnecessary with this collection. The authors are all, more or less, in agreement that a) comparison has a deserved bad reputation and we have spent some time running away from it as a practice, and b) that this has only rendered the process of comparison opaque and we must creatively rework how comparison is enacted. This is addressed by all of the authors without any grandstanding or theoretical hi-jinkery. The overwhelming sense is of sleeves rolled up and hard work being done systematically.

The stakes of rejecting comparison are made very clear by Rijke et al. In their discussion of studies around ranking and metrics in Dutch Healthcare and Biomedical Research, they demonstrate the performativity of measurement by seeking to view one disciplinary practice through the lens of the other. In doing so, they deliberately enact that well-known academic problem of being judged by the standards of another (i.e. social science evaluated by the income-generating metrics of engineering) in order to expose unexpected connections and contrasts. One tension they reveal is between the 'technologies' of comparison, when they describe how, after studying the way scientists and medics struggle to redefine one another's metrics, the research team retreats to a coffee shop to do their own comparative work. But this is rapidly counterbalanced when the authors note their participation in high-level benchmarking and performance management activities. Science may have the numbers, but the social scientists seem to have the discursive resources to do comparison differently.

Another risk is that of being captured in large research initiatives, such as EC framework programmes, that require comparison as part of the project of constituting a common European research agenda. Three chapters are dedicated to reflections on such experiences and the difficulties of working with comparison (this will do little, I suspect, for the mood or sympathy of those readers who would only just stop short of killing to secure this kind of funding). Akrich & Rabeharisoa offer an 'auto-ethnography' of their particular FP7 project. They describe, using material from meetings and emails, how the project team found its way through the variety of comparisons that they had implied in their initial proposal. In the end, delaying the 'theoretical moment' after rather than before the comparison proved the way forward.

Gad & Jensen worry in a similar vein about the demand for comparison and where it comes from. They point to 'indigenous comparisons' as a means of analysis. Comparisons are a part of the field rather than simply imposed by the analyst, and they may well be embedded into technical arrangements that do a work of auto-comparison. What is particularly interesting here is the resonance with ethnomethodological concerns about 'members categories' and the place of the analyst in relation to these 'bottom up' features of (techno)sociality. As with Rijke et al., part of the message here seems to be that our discursive skills at forging comparison are themselves part of a complex comparative relation with the language games of those we study. This is demonstrated well by Meyer who, in effect, offers a form of Membership Categorisation Analysis to show how the descriptive practices of biohackers have implications for their identity work.

Although there are moments throughout the collection where there is a sense that the thematic of comparison has led to the recuperation of existing epistemic procedures, there a plenty of moments of genuine novelty. Gad & Jensen's argument for 'lateral comparisons' is highly suggestive for the mobilisation of ethnographic description in STS. Similarly, Lutz's notion of 'comparative tinkering' is far more promising than the now standard recourse to the term 'experimentation' to describe the creative labour of reframing data through loose and partial connections. Faria beautifully illustrates how 'comparing the incomparable' (in this case different examples of colonial architecture in India) can be done thoughtfully and rigorously. Just don't try convincing historians of that.

Some of these threads are drawn together in Deville et al.'s chapter, which places the collectivity of analysis at its centre. They describe their research team as having become a 'comparator' – a device that produces auto-comparisons when 'fed' with relevant data. In doing so they re-introduce an important sense that our analytic skill does not come from comparing anything with anything, but rather from the situated act of being exposed to particular opportunities where we work in concert with one another to develop and exploit intellectual and practical connections. This takes us back to Akrich & Rabeharisoa's notion that doing research places us into the trap of comparison, but also delivers us the access and data to find a way out.

Isabelle Stengers (2011) haunts much of this collection, particularly in Robinson's concluding chapter, which ends up effectively saying that the reader should take a look at Stengers' 'Comparison as a matter of concern' piece. I very much doubt that many readers of this volume will not already be familiar with that work. But what they will find here is a fine collection of essays by researchers who engage with comparison as a live analytic and empirical concern, and who have the authorial skills to convince that this an essential debate with profound implications.

References

Viveiros de Castro E (2004) Perspectival Anthropology and the Method of Controlled Equivocation. *Tipití: Journal of the Society for the Anthropology of Lowland South America* 2(1): Article 1.

Stengers I (2011) Comparison as a matter of concern. *Common Knowledge* 17(1): 48-6.

John Law & Evelyn Ruppert (eds) (forthcoming, 2016) Modes of Knowing: Resources from the Baroque. Manchester: Mattering Press.
ISBN 978-0-9931449-8-1.

Benjamin Alberti
balberti@framingham.edu

A Fifth Mode of Knowing

The challenge for *Modes of Knowing: Resources from the Baroque* is this: how can "cultivating a baroque sensibility" (Verran & Winthereik) reveal and hold onto the messy and otherworldly experiences we encounter in research without, simultaneously, smothering them with an excess of words? The goal of a particularly baroque mode of knowing (the fifth of five such modes that John Law quickly sketches at the opening of his introductory chapter) is to go beyond the boundaries of the academic in answering the question, "How should we know the world?" How do we "get at the richness of experience that escapes social science representation" (Verran & Winthereik)? Most chapters in the volume (all by established figures) struggle faithfully with what this might mean for their research projects. Mobilizing heterogeneous materials and inspired by baroque and contemporary art, music, architecture, food, stones, Indigenous concepts and numbers the authors work to release social science research from its own constraints. To this end, the volume is a success, though, inevitably, questions linger.

Coherence among chapters is strong, with many references to Law's introduction (in which he lays out the characteristics of a baroque mode of knowing through reflections on Bernini's sculptures). The volume is firmly positioned at the leading edge of the methodological concerns of STS (though certainly not limited to this field), and is characteristic of Law's (e.g., 2004) work. I am enthusiastic about this project, which follows the creed that reality is enacted and knowing, therefore, is embedded in practices. As Evelyn Ruppert puts it: "Methods do not stand apart as representations of social worlds; they also perform those social worlds."

Even so, the nature of the task suggests a contrivance. Is "baroque" simply a faddish way to describe what we are already doing? As van de Port reminds us, Hellen Hills (2011: 31) in *Rethinking the Baroque* warns that the concept is readily extendable and risks becoming meaningless. However staged the effort, the charge to "think baroque" has produced some wonderful effects, principally for two reasons. First, the deliberate staging turns out to be important; and second, the materials that are engaged with, whether "of the Baroque" or "baroquely," reveal a material excess that pushes and prods the authors to explore "different kinds of realities" (Law).

Mining the baroque for concepts, Law provides a list of baroque techniques of knowing. I did shuffle a little in my seat at the happy coincidence that several of these terms, and others used frequently, seemed delightfully close to those of our most cherished contemporary theorists, such as the ease with which Deleuzian concepts resonate with Law's Baroque. MacKenzie's chapter on statistical probability and big data, for example, where he convincingly makes the case for post-demographic probability in which individuals themselves are distributive numbers, or "modes of the world," was unsurprisingly close to a Deleuzian reading of Leibniz' monad. Similarly, the concept of "performativity" featured frequently. Hennion's chapter used it to explain the nineteenth-century reinvention of J.S. Bach's music (and public), which led to fierce debates about authenticity. Baroque music exists only in the reiterated playing. Physicist Karen Barad appears on the fringes of several papers; her mechanistic world-formulae seeming strangely at odds with a baroque orientation (though Blaser's summary of Barad's diffractive method is exceptionally good).

Curiously, Raffles' chapter makes no mention of baroque. Either, generously, Raffles simply distills a baroque sensibility or, less generously, one could identify his work as Baroque if one wanted to. That is, one needn't. What is it, then, that a Baroque mode of knowing has to offer that other approaches don't? The concept of "excess" appears frequently in the volume, and I think is a good part of its success. It is used as a mark of the ineffable, the element of experience beyond the reach of conventional research. Van de Port, for example, is concerned with the perennial ethnographic question of how to reconcile the excesses of fieldwork with the narrative style of traditional anthropological accounts. Baroque is open to representational modes that get at the "rest of what is" (excess). Blaser shows most clearly how excess can be generative, diffracting it through the Yshiro concept of the *yrmo* as an "illuminating shadow" to analyze the incommensurability of understandings of what is going on during community workshops on conservation.

The excess that I find most compelling lies in the incredible array of things, materials, colors, sounds, forms (folds), artworks, and sensualities; it is materials first almost always. Verran and Winthereik's particularly lucid chapter presents a contrast between the technoscience engineering object (exemplified by a diagram presenting the potential steps towards promoting wave energy innovation) and a sixteenth century baldachin (a tapestry throne canopy). Both are diagrams, devices that are "ephemeral clots of material semiotic resources" with the dual function of representation and "pilotage" (material guides towards future action). The technoscience diagram represents but does not guide; its aim is non-contradiction. The baldachin does both – materially foregrounding "complexity, openness, and emergence" and enabling equivocation. Hugh Raffles' in search of the London Stone is guided by multiple traces of its material life. The Stone in fact "tangles" time. Its slow, oolitic life of geological time and hard materiality would appear to anchor it – but it is repeatedly cut adrift by Raffles' fractured description until it more closely resembles a fluid vortex. And what would appear to be a contrast, therefore, with the clafoutis – the sugary fruit dessert of Mol's wonderfully evocative chapter – turns out to be none at all. Mol argues that unlike a baroque church (or the London stone, say) in which the parts are permanently held together, the baroque coherence of the clafoutis (a composite figure composed of the absence/presence of diverse worlds) is just as resilient.

Prodded by the material excess of artworks, neat interpretations are visually unsettled. Descriptions of Bernini's sculptures are used to great effect by Law to draw out the conceptual apparatus of a baroque mode of knowing. Ruppert, in turn, uses the (x)trees project by artist Agnes Chavez – in which subjects can introduce text via data visualization technologies – to suggest how enactive research can incorporate the experiences of it by its subjects. But this material excess also reveals a problem. In his study of *lapinhas*, montages of Our Child Jesus on the Mountain, and baroque churches in Bahia, van de Port argues we should experience a baroque aesthetics rather than grasp its meaning. The challenge is how to get this into our academic texts, as he points out. Such attempts always enact cuts: Bernini stands in for Baroque in general; the description and images of the Igreja de São Francisco in Bahia limit what is essentially an overwhelming experience, a "golden storm."

A baroque coherence is undoubtedly achieved by the book, though questions remain (as they should – baroque points towards "apertures" not closures). For example, what would happen if such foundational STS concepts as performativity were themselves subjected to the full force of the baroque material? And uneasiness or a self-consciousness can be detected among some authors about the book's staging. But Baroque performances were lavish theatrical productions, very deliberately contrived, and ones that you could not easily escape, "to see it all you are forced to step inside the theatre," as Law reminds us. And this is surely part of exactly *this* project.

References

Hills H (ed) (2011) *Rethinking the Baroque*. Farnham: Ashgate Publishing Limited.

Law J (2004) *After Method*. London & New York: Routledge.

Vicki MacKnight (forthcoming, 2016) Imagining Classrooms: Stories of children, teaching and ethnography. Manchester: Mattering Press. 198 pages. ISBN: 978-0-9931449-6-7.

Josefine Raasch
josefine.raasch@rub.de

The emergence of research topics; methods and methodology; imaginations and imagining; taking perspectives and theorizing: focusing on the workings of classrooms in a range of schools in Melbourne, Australia, MacKnight's book unfolds as lively, crisp and thrilling analysis. Written for readers interested probing distinctions between analytics derived from foundational metaphysics and analytics expressing relational metaphysics, the book concerns children and their imagining, classrooms and schools; and different ways of doing imagination. Throughout the book we get to know students, teachers and schools, and the ethnographer/the writer herself.

Researchers keen to discover new and ethical ways of coming to conclusions based on a vast amount of research data, should read this book. Engaging the reader with various theoretical approaches and their implications, MacKnight accounts 'the how' of interpreting data in processes of theorizing. I will highlight two aspects of her book: its theoretical discussion of classrooms, and the ways the author constitutes accountability in the research results. But there is much more in the book to divert and delight the reader.

Drawing explicitly on approaches of Donna Haraway, Ian Hacking, Annemarie Mol and John Law, Andrew Pickering and Susan Leigh Star, MacKnight applies STS methods to educational settings, and in contrasting foundational and relational analytics the book owes much to Helen Verran's work. In line with Ian Hacking, MacKnight (2016: 20) suggests that 'good' knowing does not refer to the best representation of a reality, but to 'what enables knowers best to intervene in the world'. Accordingly, knowers come to know the real by their own interactions with it. Consistently recognizing and dealing with multiplicity, MacKnight introduces a second approach to what is 'good' by referring to Kathryn Pyne Addelson (1994: 1), who has 'the good' emerging in relations. Indeed, what place could be better to see knowers emerge in relational situations than classrooms? But what *is* a classroom? It is an unusual question to start with, yet as MacKnight delves into this consideration, she situates her research locally, socially and theoretically.

MacKnight (2016: 37) suggests two modes of ordering for exploring what a classroom is: A classroom could be ordered as part of a whole, embedded into a school that itself is embedded into the society, and/or as a complex assemblage of classed and classing bodies. If classrooms are ordered as parts of a whole, nested within society, how, then, are they related to that society? To which educational traditions, for example, do the schools subscribe? What values do they represent? By referring to Kieran Egan's (1997: chapter one) work, particularly his three educational aims, MacKnight (2016: 39) initially follows this path of taking schools as representative of values of an ideal society, until she concludes that it becomes problematic if schools represent more than one conflicting ideal value. Rather than asking how schools represent values, the author suggests that we think of *doing* these values. Focusing on practices of doing values enables us to become more able to deal with multiplicity (MacKnight, 2016: 40). Thus, she suggests a change of perspective, which looks at classrooms as assemblages of classed and classing bodies.

Paying attention to the metaphors in the classroom, and also to metaphors for the classroom, showing how they work, MacKnight argues that the metaphors guide our (re-)constituting of our world. Here, MacKnight (2016: 47) is drawn to the metaphor 'class' that is 'both metaphysically and materially significant to what classrooms are'. The terms 'classed' and 'classing' are used to accentuate two features of the people assembled in a school: a) they are embedded in hierarchies of wealth, culture, and future potential and b) 'they are bodies who learn to classify their worlds in particular by everyday ways' (MacKnight, 1916: 37-38). Having understood them this way, MacKnight shifts to the issue of classing/classifying and making groups, group identities and memberships. In this assemblage, learning to class/classify becomes one of the key tasks of primary schools. Each of the school types has different aims in terms of what the children should ideally become, but rather than analyzing through those distinctions, the schools educate the author to link them in their socio-material way of doing imagination. She asks which patterns of thinking were encouraged in the classrooms, as well as how mental connections and separations were made explicit by students and teacher. The roles of difference and sameness in analysis are vividly and explicitly demonstrated here. Unusually here we see sameness mobilized explicitly within the articulation of difference. This makes the analysis revealing in a novel way.

Each chapter and each subchapter starts with field notes. From there, MacKnight shifts refreshingly between analysis, the metaphysical implications of this analysis and reflections on her process of writing. It is difficult to do justice to the elegance of MacKnight's argumentation. However, a glimmer of the creativity, the careful selection of and thorough dealing with theories and their implications should be mentioned here. I followed each of MacKnight's questions and her attempts to tackle them with increasing curiosity. Indeed, they allowed me to open up to the described 'possibilities to interpret imagination as the routines of perspective taking' (MacKnight, 2016: 143) in my own analyses.

Similarly remarkable was her 'working up' of the empirical data. Being at the office again after three months of fieldwork, MacKnight faces the task that is familiar to most of her readers: analyzing an enormous amount of research data. Ten thousand words of field notes, hundreds of paintings and stories created by children, three and a half hour of interviews, school brochures and curricula have to be ordered in one way or the other. "Well, I thought, nothing will happen if I just sit here; I should *do* something" (MacKnight, 2016: 60). So, she starts piling, un-piling and re-piling the paintings of the children, thinking about the arguments she could make with each of the piles and the theoretical formats she is adopting by doing so. MacKnight makes visible the ethical and political consequences of ordering the data according to 'generalization' and 'typologies', 'meaning' and 'the politics of becoming'. She makes us see how the selected approaches makes more familiar or less familiar sense, arguing that the politics of experimentation, enacted by theorizing imagination as the unpredictable outcome of human and non-human interactions, is the most ethical approach. As with the elegant line of argument, the section of the book on experimental theorizing convinces the reader with a seemingly effortless narration, guided by careful reflection.

MacKnight makes the relations that constitute the entities of her research: the classroom, the students, the teachers, and the imagination. The strength of this book is the consistent tailoring of a relational writing about generative practices while applying relational metaphysics. This book makes me a modest witness of MacKnight's thinking; her thinking *with* implications of data collection methods, and theorizing. I felt stimulated and encouraged to experiment with perspective-taking, to follow the data, and to get engaged in the knowledge process.

References

Egan K (1997) *The Educated Mind: How Cognitive Tools Shape our Understanding.* Chicago & London: University of Chicago Press.

Addelson K P (1994) *Moral Passages: Towards a Collectivist Moral Theory.* New York: Routledge.

Franck Cochoy (forthcoming, 2016) On Curiosity: The Art of Market Seduction. Manchester: Mattering Press. 138 pages.
ISBN: 978-0-9931449-4-3.

Gay Hawkins
g.hawkins@westernsydney.edu.au

Striptease, packaging, click bait, shop windows. What do these disparate things have in common? They all work on the assumption that an interested subject wants to know and see more. Or, to put it more simply, they all provoke curiosity. Devices to activate curiosity are everywhere around us, and many of our modes of being in the world are driven by curiosity yet this phenomenon remains strangely unexamined. It is not until you read Franck Cochoy's wonderful book *On Curiosity* (one of the first published by the exciting new open access Mattering Press) that you realise how little we know about curiosity, and how central it is to so many ordinary practices. From the opening pages Cochoy makes us curious about curiosity. The book has a compelling intellectual energy, you can feel Cochoy's desire to know more about curiosity, to 'unpack' it, and this energy is contagious. An author wanting to know more … produces a book … that provokes a reader … who wants to know more: this is the force of curiosity at work. Cochoy is not ashamed about his curiosity about curiosity, instead he celebrates it as fundamental to the intellectual disposition. And what a refreshing admission this is.

The explicit project of the book is to develop a history and sociology of curiosity through the example of the market. As one of France's leading economic sociologists working in the STS/ANT tradition Cochoy has produced an impressive and influential body of research that has consistently foregrounded the role of devices in organising markets. From packaging to the shopping trolley to data matrixes Cochoy has examined how markets are agenced by sociotechnical artefacts. In this study these and many other devices are re-examined with the aim of understanding exactly how they work as 'captation devices', or tools for equipping relationships between organisations and their users or audiences. Captation devices are not designed to manipulate rather to *capture*. That is, they ascribe an attitude to people that is familiar or that they are willing to give attention to, and they suggest a possible mode of action that they may not necessarily have had in mind. The key attitude or disposition that Cochoy examines is curiosity: what particular devices activate it and allow it to be expressed throughout society? And what types of actions does it prompt?

Before getting to the issue of markets as the heartland of curiosity, as the site where curiosity finds its 'natural' home and contemporary commercial realisation, Cochoy takes us on a journey through the intellectual history of curiosity. This is not a review of the existing literature. For a start there is surprisingly little but, more significantly, Cochoy is far too creative a thinker to use that worn out analytic device. Rather, it is a rich exploration of the complex cultural work that expressions of curiosity have performed over centuries with the aim of understanding how curiosity becomes so central to the rise of market societies. Using a detailed analysis of the tale of *Bluebeard*, a fairy story about a wife who was unable to resist the lure of curiosity, Cochoy analyses how this story works as a profane version of earlier religious and mythical accounts of the moral dangers of curiosity. With great skill and precision the reader is taken through the various processes whereby curiosity is secularised. While religion and myth framed curiosity as both inherent to the human condition and also dangerous, the enlightenment relocated it within the realms of organised knowledge and the positive desire to know the world. Science legitimated curiosity but also disciplined it, literally and metaphorically. Public experiments, the rise of observation and fieldwork, natural history displays all worked to activate curiosity and also regulate it. In many senses what Cochoy is often describing is the governmentalisation of curiosity, the processes whereby it becomes incorporated into the management of populations and the conduct of conduct, as Foucault terms it. In the rise of the museum, we don't simply see the displacement of cabinets of curiosity with virtuous curiosity driven by the quest for knowledge for its own sake. We also see how looking at a collection expresses a pedagogic imperative; the desire to improve specific populations through the activation of managed curiosity.

The marketization of curiosity is Cochoy's primary focus. Over three chapters a series of devices and arrangements central to the rise of the modern commercial economy are examined. Window displays, packaging, advertising, data matrixes and more are analysed in terms of exactly how they made curiosity benefit markets. These analyses are richly historical and empirical. They also draw on numerous disciplines from sociology, to philosophy to anthropology. It is impossible to do justice to the complexity of these examples except to say that they foreground both the specificity and multiplicity of ways in which curiosity can be activated within markets – as any device centred analysis should. So, for example, in the analysis of the rise of window displays phenomenological accounts of the meanings of doors, locks and mirrors are drawn on to show how curiosity is activated in a looking subject when they encounter the shop window. And how this activation pulls the subject into a world of objects and a distinct form of consumer captation. Other examples examine the relationship between packaging and surprise, the role of advertising billboards in making curiosity a generative public event and the ways in which data matrixes on goods only grant access to the knowledge they possess when they are activated.

What links these cases and gives them coherence is a thoroughly STS commitment to understanding how technologies prompt curiosity, how they make it real. Cochoy is interested in how curiosity is done or performed in market relations and that means refusing to see it as an internal expression of subjectivity. One of the most powerful and convincing lines of thought in this book is the insistence that curiosity is not a human motive or action that is spontaneously available. It has to be aroused or activated and not only is this process distributed across a myriad of market relations but the techniques for arousing it shape its expression. This argument disrupts neat distinctions between market dispositifs and consumer dispositions. Instead, Cochoy argues, the work of qualification works *both* ways: it qualifies goods and their social spaces *and* people by offering them motives and actions that justify becoming attached to goods; that realise specific types of consumer curiosity.

There is much more to say about the complexities of curiosity and market seduction set out in this book but let me finish with a few points about its significance for wider debates within STS. I read *On Curiosity* soon after finishing Muniesa's (2014) *The Provoked Economy*. Inevitably, both books entered into a conversation. This conversation focussed on the similarities and differences between the terms: 'provoked' and 'activated'. 'Activated' is a term Cochoy uses a lot, along with aroused, cultivate and awakened, it foregrounds the dynamic ways in which markets are animated or realised. In his account curiosity works as a force that is not grounded in either buyers or sellers rather, it is an effect that must be enacted to make markets work. Consumers must identify it as a motive for action and sellers must find the right devices to awaken or call it up. The activation of curiosity is generative of market relations and dispositifs. For Muniesa, 'provocation' refers to a performative take on reality or the ways in which the real is realised as an effect of various provocations. Technology or devices or experiments provoke effects, they trigger latent energies of everyday life or events and in doing so reveal a new reality. In Muniesa's performative and pragmatist approach categories like consumers, motives, calculations or supply and demand do not precede markets they are an effect of them. Provocation then, is thoroughly compatible with Cochoy's notion of activation in the sense that both concepts frame the social as effectuation or enactment.

However, what is very valuable about Cochoy's account of curiosity is the way in which it foregrounds the *activity* of activation. All through this book I kept wondering – is curiosity an emotion, a disposition, a passion or what? Cochoy explores all these options. Finally, I understood that it was a force or form of distributed agency that helps make markets. As Cochoy says in the concluding pages 'In spite of all the religious and moral obstacles that have stood in its way, curiosity therefore truly remains a force of action.' While it might seem like an innate human possession the impetus to know, to break out of habits, to investigate the new or explore change has to be aroused and ordered, and in this activation curiosity is both constituted, acquires agency and prompts actions.

One of the longstanding critiques of STS approaches to markets is their failure to develop a critical edge, to account for power and domination. These claims are part of a long debate that I don't wish to revisit. Rather, I want to point to Cochoy's account of curiosity as offering an exemplary way of understanding market powers in a non reductionist way. If we see power not as a noun or possession but as an activity, as a capacity to shape performances in certain ways, then it is possible to understand the significance of curiosity as a force that activates and animates markets and shapes their effects. Cochoy's achievement in this superb book is to show how the force of curiosity has been historically anchored in markets, helping them realise or constitute a multiplicity of realities not just economic ones. His project is not to judge the effects of curiosity but to understand exactly how this force effects or constitutes market realities. This, to my mind, is far more politically valuable than critique for it points to the contingencies of markets and their constant vulnerability to contestation. And, more significantly, to the activation of curiosities that can make trouble for markets and provoke significant political effects. *On Curiosity* is a great book – thinking like this reminds us of how powerful 'the market' is for investigating so many aspects of the social.

Reference

Muniesa F (2014) *The Provoked Economy.* Routledge: London.

Science & Technology Studies

Volume 29, Issue 2, 2016