Science & Technology Studies

4/2016

Science & Technology Studies

ISSN 2243-4690

Subscriptions

Subscriptions and enquiries about back issues should be addressed to:

Email: johanna.hokka@uta.fi

The subscription rates (2016) for access to the electronic journal is 40 euros for individual subscribers and 100 euros for institutional subscribers.

Science & Technology Studies

Volume 29, Issue 4, 2016

Guest editorial

Articles

Reviews

Visit our web-site at

www.sciencetechnologystudies.org

Knowledge Infrastructures: Part IV

Helena Karasti

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

This issue of Science and Technology Studies is the final one of four in total published this year focusing on the topic of Knowledge Infrastructures. Across the four issues we have presented fourteen papers (thirteen research articles and one discussion paper) and four book reviews. In this final editorial we first take a look at the issues raised by the final batch of articles, then take a step back to review the collection as a whole, considering what it tells us about the state of the art in Science and Technology Studies' understanding of knowledge infrastructures and looking forward to the challenges still on the horizon.

Articles in This Fourth and Last Part of the Special Issue

The first article 'The Daily Shaping of State Transparency: Standards, Machine-Readability and the Configuration of Open Government Data Policies' addresses the issue of open standards for diffusing online data in the context of government bureaucracies. In common with open data initiatives in other substantive fields, such as science (Borgman, 2007) and cultural heritage (Stuedahl et al., in this issue), many governments are now committed to the release of open data. Open Government Data (OGD) initiatives are constructing ways to store and share data, forming a new layer of 'open data infrastructure' shaped by the development and deployment of data standards (Lampland and Star, 2009). While OGD movements to sharing data under non-proprietary standardized formats have been highly visible, Samuel Goëta and Tim Davies point out that considerably less attention has been given to what is happening on the ground around the production of standards and the actual consequences of standards for knowledge workers, the issues that form the authors' focus in this article.

Goëta and Davies study three very different open data standards, namely Comma Separated Value (CSV), General Transit Format Specification (GTFS) and the International Aid Transparency Initiative (IATI). They operate an 'infrastructural inversion' by looking at the historical development of the named standards and by studying ethnographically the 'back rooms' of government bureaucracies with a focus on the invisible work necessary to open the data by using these standards. The authors pay particular attention to the concrete work practices that go along with aligning the standards, the organizational arrangements they create, and the way they shape the data for others to access and use.

Through the empirical work the authors discuss how transparency is or is not achieved by the demands for openness and standardization. The authors show that the standards substantively shape the production of open data. They describe how the use of open standards requires intensive work in order to transform and adjust datasets to the standards; thus, the making of datasets machine-readable may increase the complexity of releasing data. The authors further show how enacting open standards operates "a quiet and localised transformation of bureaucracies", with consequences for how open government data and transparency agendas are performed. The use of open standards has become interpreted not only as a sign of a quality dataset, but also used to evaluate the progress of the open data program itself. The adoption of open standards is increasingly becoming (used as) an indicator of the advancement of open data programmes. Furthermore, the authors discuss the particular kind of transparency delivered by OGD which reveals a rationalisation and representation of the information held inside the state, focussing on machine-mediated transparency rather than transparency as a relationship between citizen and account-giving state.

In addition to the above 'producer' side inside the 'back rooms' of government bureaucracies, the authors also discuss the 'user' side of OGD. They see that the emphasis on machine-readability in OGD projects configures the primary users as 'advanced users' with a need for technical skills, financing and capability to create services to make desired re-use of the published data. These set-ups (of professional developers and ecosystems) introduce other layers of infrastructure and eventually intermediation between citizens and the state.

In the second article, Ayelet Shavit and Yael Silver discuss the development of long term biodiversity surveys and specifically focus in on tensions inherent in recording locality within such surveys. The first case study in the article discusses the evolving treatment of locality information within the specimen collections of the Museum of Vertebrate Zoology at the University of California, Berkeley. A formalized approach to recording was established early on in the museum's history, requiring both a standardized set of information including a record of locality and a narrative account of the circumstances surrounding collection of the specimen in a field journal. This system of recording thus combined what Shavit and Silver term 'exogenous' and 'interactionist' approaches to locality. The two approaches are associated with contrasting epistemic values: an exogenous approach to 'location' focuses on production of representative and reliable data whilst the interactionist approach attends to the need for comprehensive and accurate data for the location in question. Both systems co-existed in the pre-computerised system of journals, index cards and tags, but the advent of computerized records in the 1970s began a push towards inclusion of a searchable and generalizable version of specimen locality in specimen databases and prompted the development of a system to map historical localities to estimated longitude and latitude using a standard georeferencing protocol. Subsequently, new challenges for the recording of locality emerged, as new devices used by researchers in the field occasioned a more precise georeferencing, producing new forms of data and shifting away from narrative field journals to numerical data. A separation emerged between the requirement for a globally interoperable and easily searchable form of locality information and the historical collections of narrative data on circumstances of collection that were locally held at the museum and mined by relatively few researchers.

A subsequent workaround involved digitization of field journals, allowing this information to be linked to specimen records and hence made available albeit not in an equivalent searchable form to the exogenous locality information.

The second case study in Shavit and Silver's paper focuses on a biological monitoring project 'Hamaarag' initially associated with Long Term Ecological Research (LTER) stations funded by Israel's Science Foundation. Shavit and Silver track the changing political, financial and scientific focus of the project over time, and also the tensions over the version of locality embedded within the project. As with the Museum of Vertebrate Zoology, tensions focused on a clash between the possibility of developing an interoperable infrastructure across the various LTERs involved and the very different demands imposed by the different species each were monitoring and the practices of the groups of scientists involved. Shavit and Silver track the diverse and shifting pressures that beset the project over time and challenge attempts to produce a single overarching infrastructure for the project, leading ultimately to an approach that favours an interactionist approach to location and includes citizen science initiatives alongside research team efforts. Across the two case studies, Shavit and Silver identify a tension between different notions of locality and an emergent recognition that to focus only on a globally interoperable exogenous version of locality may entail a loss of a significant flexibility. They conclude that developing an infrastructure to sustain local memories of a locality and alternating between local and global memory practices (Bowker, 2005) may be better justified, both rationally and sometimes morally. Tracking the movement from a technical thing (the technical category of 'location') becoming a problematic epistemic thing, the article demonstrates a recurring issue in knowledge infrastructure work more broadly i.e. the weight that may be carried by technical decisions on the representation of key concepts.

The third article in the special issue, by Dagny Stuedahl, Mari Runardotter and Christina Mörtberg, focuses on the substantive field of the cultural heritage sector. The authors develop two case studies of digital infrastructure projects that are involved in opening up cultural heritage institutions to engagement with the public. Whilst both projects are working within an environment that encourages openness and public involvement, the two case studies contrast significantly in their institutional form and in the approach they take to defining what will count as an acceptable open engagement with the public. The first study focuses on a "top-down" initiative in the design phase: a new infrastructure intended to facilitate public access to archival materials. By studying discussions in the design phase Stuedahl et al. are able to identify tensions and controversies around the implementation of the high-level policy imperative to open data and engage with citizens. When these imperatives meet with local practices they encounter considerable concerns that revolve around the extent of openness deemed desirable and the quality of content acquired through crowd-sourcing, leading ultimately to adoption of an approach focused on providing access to existing archival data rather than acquiring new data. The second case study explores a 'bottom-up' initiative: a local history wiki used by professional and amateur local historians. Here Stuedahl et al. encounter the project when it is already up and running, and analyse threads from the discussion forum that demonstrate ongoing negotiations over the categories to be used to structure contributions to the wiki and tensions between wiki administrators and local historians over the extent to which diverse understandings can be accommodated within the wiki.

To draw together the comparison between these two substantively similar yet contrasting initiatives Stuedahl et al. rely on the concept of 'attachments' used within STS variously by Gomart and Hennion (1999), Latour (1999), Marres (2007) and Hennion (2012) to denote an array of resources that are drawn on to inform and make sense of engagements and actions. Attachments are potentially more diffuse than motivations and more emotionally charged than influences, offering a means to identify what matters to people as they decide on a course of action or design an intervention. In the participatory knowledge infrastructures that they study Stuedahl et al. identify attachments used by actors to outline what matters to them and position themselves in relation to past, present, and future. The authors argue that attachments offer a useful alternative way to explore the temporality of knowledge infrastructuring, stressing that sustainable infrastructures may need not only to work with the long now (Ribes & Finholt, 2009) of an anticipated future but also to display an appropriate attachment to relevant values and practices of the past as well as attachments to other pressures and policies in the present. By highlighting the various attachments that actors bring to the two case studies they outline, Stuedahl et al. bring out the process through which the contrasting (and sometimes internally conflicting) notions of openness and engagement that the two projects arrive at come into being.

An Overview and Emerging Themes

The fourteen articles published in this special issue, while all viewing their material through the lens of the knowledge infrastructure, have covered a range of substantive fields: biodiversity (Taber, 2016); cultural heritage (Stuedahl et al., in this issue); disease genetics (Dagiral & Peerbaye, 2016); drug discovery (Fukushima, 2016); e-health (Aspria et al., 2016); ecological science (Stuedeahl et al., 2016; Shavit & Silver, in this issue); environmental monitoring (Jalbert, 2016; Parmiggiani & Monteiro, 2016); open government (Goëta & Davies, in this issue); public health (Boyce, 2016); social science data archiving (Shankar et al., 2016); weather recording (Goëta & Davies, 2016); wikipedia content (Wyatt et al., 2016). While many have at their heart a database or other form of digital technology, this has not been universally the case: Taber (2016) views the herbarium as the focus of a knowledge infrastructure. The articles exemplify the interdisciplinary trend within Science and Technology Studies more broadly. While we have not conducted a systematic census of the disciplinary origins of the scholars represented here, it is clear from their institutional addresses as much as their substantive foci that the authors come from an array of backgrounds including anthropology, informatics and information science, media and communications, public health and social science in addition to science and technology studies departments. The geographical spread is also broad, including authors from Australia, France, Ireland, Israel, Japan, Netherlands, Norway, Sweden, United States of America and United Kingdom.

In the three previous editorials (Karasti et al., 2016a, 2016b, 2016c) we have identified some emerging themes that tie together the contributions made by individual articles and suggest areas of common significance across quite diverse manifestations of knowledge infrastructures. In the first issue we discussed themes of scale, invisibility, tensions, uncertainty, and accountability. We also explored methodological issues, focusing on the infrastructural inversion and the challenges inherent for the researcher in choosing levels, locations, and scales to examine. In the second issue we explored the performativity of knowledge infrastructures and the struggles over power, values, and voice that prevail at the very heart of infrastructural work. The third issue highlighted temporality and labour as key areas of connection across infrastructural studies.

These themes continue to resonate across the three articles presented in this fourth issue to focus on knowledge infrastructures. All three articles deploy a methodological focus that encompasses the diverse scales of infrastructural work and each in its own way highlights an otherwise invisible or neglected aspect of that work and brings it into the foreground as consequential site for the enactment of values and the experience of tensions between different practices and sets of accountability. Temporality arises with particular significance in Stuedahl et al.'s exploration of the notion of attachments, as they argue that an attachment to aspects of the past can give meaning to infrastructural work as much as visions of an anticipated future.

Beyond the themes already identified, a further theme deserves exploration in this editorial: the notion of openness. As a value and a set of practices the notion of openness has a considerable contemporary significance and yet, as studied here, it emerges as a problematic concept not necessarily easy to achieve. Openness appears repeatedly across the papers collected here: in the first issue, Parmiggiani and Monteiro (2016) explore the development of an infrastructure for monitoring subsea ecosystems and evaluating environmental risk and here achieving a portrayal of the openness of data in a public portal plays a part in building a new sense of trust; in the second issue, Shankar et al. (2016) propose a study of social science data archives that pays attention to the specificity of circumstances under which open sharing of data arises; in the third issue Aspria et al. (2016) explore the metaphors that underpin operationalization of a patient information portal that aspires to be seen as open and inclusive. In this fourth issue, openness receives further significant attention: Goëta and Davies place the standards that underpin open data sharing under the spotlight, and find that these standards are a site of considerable labour both in development and use and far from a smooth route to automatic transparency; Stuedahl et al. focus on the movement towards open data sharing in cultural heritage contexts and find that whilst aspiring to openness may be dictated by policy, it still requires considerable negotiation to make manageable in practice. When we study contemporary knowledge infrastructures we find values of openness often embedded there, but translating the values of openness into the design of infrastructures and the practices of infrastructuring is a complex and contingent process.

In putting together the special issue we aimed to assess the current state of Science and Technology Studies' contribution to the understanding of knowledge infrastructures. This set of emergent themes, connecting across together, exemplify the contribution that a set of sensibilities drawn from Science and Technology Studies can make in this area: by a detailed attention to technology as it is enacted in situ and as it is embedded in and embeds policies and practices, we can see the knowledge infrastructure as a very particular kind of achievement with far-reaching yet often overlooked consequences. We learn in detail about the modes of governance that depend upon and are enabled by knowledge infrastructures and we find out how great the gulf may be between an aspiration in the domain of policy and its realisation on the ground. STS scholars are studying the processes of infrastructuring in detail but also considering the consequences: what kind of ways of being in the world do knowledge infrastructures enable, to whom do they give voice and who do they silence, what do they prioritise and what do they neglect or negate?

Viewed as a whole, this collection of papers suggests that the STS-enabled study of knowledge infrastructures is on increasingly solid theoretical and methodological ground. Across the papers we see a confidence in identifying diverse sets of technological developments as knowledge infrastructures and applying to them a relatively stable set of theoretical resources. Among the papers we find also theoretical innovations, such as Fukushima's (2016) recourse to a Marxist-inflected notion of infrastructure alongside the resources of STS or Stuedahl et al.'s (in this issue) deployment of attachments as a means to uncover the meanings that pervade infrastructural work. On the whole, however, the articles wear their theoretical development relatively lightly and concentrate on illuminating what is being achieved through the medium of knowledge infrastructural work and how this is being brought about.

Methodologically speaking, also, this collection of papers speaks to a relatively confident set of resources being deployed to good effect. Most of the papers make a broad claim to ethnographic approaches, with the notable exceptions of Wyatt et al. (2016) in their study of data from editorial discussions on Wikipedia and Taber (2016) and Shankar et al. (2016) with historical approaches founded on archival data. Ethnography, in the knowledge infrastructure context, often means a foundation of participant observation within a key location, taking part in ongoing discussions and attending meetings. The temporal and spatial complexity of infrastructural work is handled through a combination of mobility from the research and recourse to programmes of interviewing and documentary analysis. Online discussions appear as sources of data that give a useful insight into day-to-day negotiations into the meaning of data, capturing as they do a level of detail often otherwise ephemeral and hard to capture when work goes on in face-to-face settings, even for an ethnographer on the spot. The increasing recourse to online discussion forums for getting infrastructural work done has, as a by-product, provided a useful set of data for STS scholars interested in how this work is done.

Studying the otherwise invisible becomes easier when this work is captured in a persistent form.

The notion of the infrastructural inversion has clearly become one of the established resources of an STS approach to knowledge infrastructures. Responding to Geof Bowker's call to make material infrastructures the central object of study (Bowker, 1994), many of the papers in this collection used the infrastructural inversion in the standard sense of a methodological sensitivity associated with making otherwise neglected things visible, as exemplified by Bowker and Star (1999). In doing so, these papers confirmed the pertinence of this methodological lens to scrutinize the interdependences between technical components and the politics of knowledge production. Three articles elaborated on the infrastructural inversion to a significant extent: Fukushima (2016) drawing out an isomorphism with the Marxist inversion of the infrastructure/superstructure relation; and both Parmiggiani and Monteiro (2016) and Dagiral and Peerbaye (2016) drawing out the use of the inversion as a resource by actors themselves.

There are, thus, promising signs for future knowledge infrastructure studies in STS, confidently adopting and developing a mature set of methodological and theoretical resources. Promising future prospects include possible pay-offs from making further use of online data and myriad digital traces left by digital work, taking on board Edwards et al.'s (2013) challenge to infrastructural studies to take more account of big data. Future studies may also do more to engage in depth with the reflexive work done by the actors in infrastructural projects, building on the recognition that concepts such as the infrastructural inversion resonate strongly with what actors themselves do. New methodological forms may yet emerge. The majority of the articles collected here represent either the work of one scholar, or a small group of scholars pooling or contrasting a small number of case studies. We see little as yet of the larger team-based and multi-sited studies that may be necessary in order to scale up knowledge infrastructure studies and more extensively explore their ramifications across time and space as Edwards et al. (2013) exhort. Similarly, while historical and archival studies promise to allow us to extend our interest in the evolution of knowledge infrastructures across greater time spans, as yet our analytic resources for conducting archival studies are relatively under-developed (Bowker, 2015). The collection of articles presented here demonstrate a healthy and vibrant field, with a clearly significant pay-off in terms of illuminating some very powerful aspects of contemporary world, yet there is clearly still further to go in developing the STS contribution in this area.

Acknowledgements

Beyond the team of four guest editors responsible for putting these special issues together there has been a considerable input from the journal Science and Technology Studies and from an array of anonymous reviewers. We are very grateful to the reviewers who have read articles so thoroughly and offered their wise and constructive advice. For the journal, Antti Silvast has taken editorial responsibility for the Knowledge Infrastructures special issue. Antti has been hugely generous with his time and knowledge of the field as he has guided and advised us across what has been, for him, a much greater commitment than initially anticipated when the response to the call for papers produced four special issues rather than a single one. Louna Hakkarainen has acted for the journal to see the issues efficiently through the production process.

References

Aspria M, de Mul M, Adams S & Bal R (2016) Of Blooming Flowers and Multiple Sockets: The Role of Metaphors in the Politics of Infrastructural Work. *Science & Technology Studies* 29(3): 68-87.

Borgman, CL (2007) *Scholarship in the digital age: Information, infrastructure, and the Internet.* Cambridge: MIT.

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

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

Bowker GC (2015) Temporality: Theorizing the Contemporary. *Cultural Anthropology.* September 24, 2015. Available at: https://culanth.org/fieldsights/723-temporality (accessed: 22.8.2016)

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.

Dagiral É & Peerbaye A (2016) Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases. *Science & Technology Studies* 29(2): 44-61.

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: Deep Blue. http://knowledgeinfrastructures.org/ (accessed: 5.11.2016)

Fukushima M (2016) Value Oscillation in Knowledge Infrastructure: Observing its Dynamic in Japan's Drug Discovery Pipeline. *Science & Technology Studies* 29(2): 7-25.

Gomart E& Hennion A (1999) A Sociology of Attachment: Music Amateurs, Drug Users. In: Law J & Hassard J (eds) *Actor Network Theory and After.* Oxford: Blackwell, 220-47.

Granjou C & Walker J (2016) Promises that Matter: Reconfiguring Ecology in the Ecotrons. Science & Technology Studies 29(3): 49- 67.

Hennion A (2012) Attachments: A Pragmatist View Of What Holds Us In: *The First European Pragmatist Conference*, Roma 19-21 September 2012.

Jalbert K (2016) Building Knowledge Infrastructures for Empowerment: A Study of Grassroots Water Monitoring Networks in the Marcellus Shale. *Science & Technology Studies* 29(2): 26-43.

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

Karasti H, Millerand F, Hine CM, & Bowker GC (2016b) Knowledge infrastructures: Part II. Science & Technology Studies 29(2): 2-6.

Karasti H, Millerand F, Hine CM, & Bowker GC (2016c) Knowledge infrastructures: Part III. Science & Technology Studies 29(3): 2-9.

Lampland M & Star SL (eds) (2009) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life.* Ithaca, NY: Cornell University Press.

Latour B (1999 [2004]) *Politics of Nature: How to Bring the Sciences into Democracy.* Cambridge MA: Harvard University Press.

Lin Y-W, Bates J & Goodale P (2016) Co-Observing the Weather, Co-Predicting the Climate: Human Factors in Building Infrastructures for Crowdsourced Data. *Science & Technology Studies* 29(3): 10-27.

Marres N (2007) The Issues Deserve More Credit: Pragmatist Contributions to the Study of Public Involvement in Controversy. *Social Studies of Science* 37:759.

Parmiggiani E & Monteiro E (2016) A Measure of 'Environmental Happiness': Infrastructuring Environmental Risk in Oil and Gas Offshore Operations. *Science & Technology Studies* 29(1): 30-51.

Ribes D & Finholt TA (2009) The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10(5): Article 5. Available at: http://aisel.aisnet.org/jais/vol10/iss5/5 (accessed: 4.11.2016)

Shankar K, Eschenfelder KR & Downey G (2016) Studying the History of Social Science Data Archives as Knowledge Infrastructure. *Science & Technology Studies* 29(2): 62-73.

Taber P (2016) Taxonomic Government: Ecuador's National Herbarium and the Institution of Biodiversity, 1986-1996. *Science & Technology Studies* 29(3): 28-48.

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

The Daily Shaping of State Transparency: Standards, Machine-Readability and the Configuration of Open Government Data Policies

Samuel Goëta

Telecom ParisTech, Social Sciences Department, France / samuel.goeta@telecom-paristech.fr

Tim Davies

Berkman Klein Center for Internet and Society, UK

Abstract

While many governments are now committed to release Open Government Data under non-proprietary standardized formats, less attention has been given to the actual consequences of these standards for knowledge workers. Unpacking the history of three open data standards (CSV, GTFS, IATI), this paper shows what is actually happening when these standards are enacted in the work practices of bureaucracies. It is built on participant-observer enquiry and interviews focussed on the back rooms of open data, and looking specifically at the invisible work necessary to construct open datasets. It shows that the adoption of open standards is increasingly becoming an indicator of the advancement of open data programmes. Enacting open standards involves much more than simple technical operations, it operates a quiet and localised transformation of bureaucracies, in which the decisions of data workers have substantive consequences for how the open government data and transparency agendas are performed.

Keywords: Open Government Data; Open Standards; Enactment; Infrastructure Studies; Data Assemblages

Introduction

> "It is time for science studies to investigate how data traverse personal, institutional, and disciplinary divides." (Edwards et al., 2011)

The case for using open standards when diffusing online data has been widely discussed for both scientific and government data (Borgman 2007; Robinson et al., 2009; Lathrop & Ruma, 2010). However, little attention has been given to the consequences of these standards for the workers involved in producing and disseminating open data, and for how standards shape the outcomes of data sharing efforts, particularly in the open government domain. Even when standards are introduced into discussions, data is often treated as though it is already available and ready-to-use, with the actual work required to construct a standardised dataset remaining almost entirely invisible (Bowker, 2000). As the proactive release of government data is increasingly presented as a "superior" mode of delivering government transparency (Birchall, 2014), it becomes vital to ask how data standards are involved in shaping government transparency? Behind the scenes, in the backrooms of open data (Goëta, 2014), what are the consequences of introducing standards for data workers and the actual organisation of government? What impact do decisions made during standardisation have upon the potential uses of open data? By understanding the challenges facing these invisible workers when working with emerging open data standards (Denis & Pontille, 2012), and the way in which standards construct practices both inside and outside the state, we can gain a deeper understanding of how an emphasis on machine-readable data comes to structure ideas and experiences of open government itself.

A growing subject in Science and Technology Studies (STS), data standards are proliferating in the development of large information infrastructures while still remaining largely invisible and taken-for-granted (Star & Ruhleder, 1996; Lampland & Star, 2009; Busch 2011). The numerous studies on open government data that have been conducted to date have largely overlooked how standards shape datasets, what they exclude, and the supplementary burden they require to be implemented. Such an approach is crucial at this particular moment, as many of the standards for an emerging open data infrastructures, embodied in data portals, policy pronouncements and common analysis and visualisation tools are currently being laid down. Rare studies have followed the information infrastructure studies program (Bowker et al., 2010) to understand open government data (Davies, 2012, 2013, 2014) but none has conducted an ethnography of infrastructure (Star, 1999) to understand the implications of these standards in the daily practices of data workers, and the consequences of these standards for the goals of open government. Situated in bureaucracies, our study aims at surfacing the invisible practical work (Suchman, 1995) that supports the implementation of open standards for government data.

In exploring emerging practices of open government data sharing, it is useful to step back to the experience of particular scientific communities over recent decades, where exchanging data has become a crucial matter and datasets are becoming an object of scientific production in their own right (Bowker, 2000; Edwards et al., 2011; Strasser, 2012). As the data required to explore phenomena of interest grows beyond that which any individual researcher or group could collect, distributed scientific collaborations have needed to develop approaches to pool and share data, leading to the creation of vocabularies, schema and markup languages for representing and exchanging data (Zimmerman, 2007, 2008). However, these processes of standardisation are not straightforward or unproblematic. Information infrastructures studies offer a rich framework within which to understand the hidden work going on in order to enable scientists to share data. Edwards (2010) uses the metaphor of "data friction" to describe the efforts required to share data between people and organizations, and it is in response to this friction that many scientific data sharing infrastructures have been developed. Yet this does not necessarily imply that the goal sought should be "frictionless data" (Pollock, 2013). Almklov (2008) finds that standardised data can be experienced by re-users as decontextualised, and difficult to extract meaning from. And several works have shown that metadata, even defined with shared and precise standards, do not lead scientists to reuse data seamlessly, as standards projects have often promised (Edwards et al., 2011; Millerand & Bowker, 2009; Zimmerman, 2008). Recognising science as essentially an open-ended and always unfinished enterprise, Edwards et al. (2011) highlight the importance of considering "metadata-as-process", and paying attention to the social negotiations that go on around data sharing in science, alongside the technical standardisation.

Open Government Data (OGD) is in many ways a younger enterprise than that of open science (Fecher & Friesike, 2014). Since the late 2000s, government across the world have been adopting policies that call for the publication of government held datasets online, in machine-readable forms, and for anyone to re-use without restriction (Yu & Robinson, 2012; Chignard, 2013; Kitchin, 2014). Multiple drivers for this have been cited, from "unlocking" the re-use value of data the state has already paid for to increasing government efficiency, and delivering greater state transparency (Zuiderwijk et al., 2012; Zuiderwijk & Janssen, 2014). As part of a transparency agenda, OGD has been discussed in relation to past regimes of reactive transparency, delivered through Right to Information (RTI) laws, which gave citizens a right to request documents from government (Fumega & Scrollini, 2011; Open Knowledge Foundation, 2011). In RTI, transparency is associated with a clear transaction between a requestor and government, but in OGD, as Peixoto (2013: 203) puts it, public actors can *"characterize transparency as a unilateral act of disclosure"*. For Peixoto (2013: 203), *"transparency may be realized without third parties scrutinizing or engaging with the disclosed information"*, although transparency theorist David Heald quotes Larsson (1998: 40–2) to argue that *"transparency extends beyond openness to embrace simplicity and comprehensibility. For example, it is possible for an organization to be open about its documents and procedures yet not be transparent to relevant audiences if the information is perceived as incoherent"* (Heald, 2006). Within the discourse of OGD, that coherence has come to be defined in terms of machine-readability, and increasingly the adoption of common open standards. OGD advocates have moved from early calls for 'raw data now' (Pollock, 2007; Berners-Lee, 2009), to argue for the adoption of open standards for data publication. Increasingly, efforts have looked to assess the success of open data initiatives with reference to these standards (Cabinet Office, 2013; Atz et al., 2015). Thus, as in scientific collaborations, OGD initiatives are turning towards the construction of new data infrastructures, shaped by the development and deployment of data standards.

Our aim here is thus to understand what is happening when these data standards are actually enacted (Law & Mol, 2008; Millerand & Bowker, 2009) in the work practices of government bureaucracies, and how this impacts upon the construction of state transparency as a component of open government. This paper is built on ethnographically informed participant-observer enquiry in the back rooms of open data: developed iteratively to look at three cases of open data standardisation: from the structuring of diverse data elements to fit with the requirements of a file format specification, through to the mapping of data from internal systems to a rich semantic standard. For each case, we attempt to operate an infrastructural inversion (Bowker, 1994; Bowker & Star, 2000) by looking first at the historical development of particular standards, the work practices that go along with aligning them, the organizational arrangements they create and the way they shape the data the public have access to, and how it can be used. Prior to introducing these standards, we first take a broader look at the role that discourses of standardisation have played in the OGD movement.

Policy and Principles of Open Government Data: Machine-Readability and Open Standards

The Open Government Data movement claims that the proactive publication of the datasets owned by public administration can lead to a new wave of innovation in the use of government data, bringing about a renewal of transparency and a transformation of administrative practices (Janssen et al., 2012). Following the launch in 2009 of the US Data.gov portal, many countries have established policy requirements and legal frameworks for open data, leading to the creation of hundreds of data portals hosting and providing meta-data on a vast spectrum of datasets, provided by national governments, municipalities, international institutions and even some corporations (Web Foundation, 2014). In 2012, G8 member countries signed up to the G8 Open Data Charter, committing to the idea that government data should be 'open by default', and including in an Annexe a list of the kinds of data, from cadastral registers to national budgets, that governments should share (G8, 2013). The G8 Charter has been followed by an International Open Data Charter (2015), which introduces a principle of data 'interoperability', and which, through its technical working group, has been exploring how to recommend data standards for governments to adopt. Within the Open Government Partnership, a voluntary association of over 60 countries committing to increase the availability of information about government activities, supporting civic participation and improve accountability, action plan commitments to open data have been amongst the most common (Khan & Foti, 2015).

Since the first articulation of common principles for OGD in Sebastopol in 2007 when well-know digital activists such as Lawrence Lessig, Tim O'Reilly, and Aaron Swartz gathered and set out eight key criteria for government data openness, machine readability and open standards have become core claims of the OGD movements. According to these principles, datasets should be provided in "machine-processable" and "non-proprietary" formats (5th and 7th principles). The Sunlight Foundation's (2010) extended "Ten Principles for Open Government Data" place a particular emphasis on the use of "commonly owned" standards, highlighting the importance of standards being freely accessible and fully documented to facilitate their use (Levien, 1998; Russell, 2014), and pointing as well to the process of control over the revision of standards, which, open standards advocates argue, should take place through a predictable, participatory, and meritocratic system (Open Stand, 2012).

This emphasis on machine-readability and open standards can be understood as a reaction against the common publication of government data either in formats such as PDF which present the layout of data, but which frustrate easy digital access to the underlying fields and figures, and the use of file formats that are protected by patents and intellectual property rights, meaning that to read the files requires either proprietary software, or paying license fees for the right and resources to decode and manipulate the data. It is also motivated by a desire to have data files which can be accessed and manipulated in as wide a range of tools as possible, such that even de-jure non-proprietary formats tend to be considered as de-facto closed by developers if established tooling for working with these formats cannot be easily found across a wide range of programming languages and software packages. However, many of the OGD portals in operation around the world still predominantly provide access to files which fail to meet key definitions of machine-readability, and, even if they do, which fail to make use of common standards (Murillo, 2014; Web Foundation, 2015), leading to redoubled efforts to promote 'best practices' for data publication (W3C, 2015). Furthermore, advocates have also been concerned with how data is represented when it is published using machine-readable open formats, looking to also see use of common schemas that define the kinds of fields and values that would be considered valid in a particular kind of data, and which tools reading that data should be able to understand.

Using open standards in releasing government data is now more than a mere principle: it is progressively being required by regulations brought in to implement OGD policies. In 2013, President Obama released a memo, which states that government information must be released under open and machine-readable standards (Obama, 2013). Agencies are required reporting progress on the implementation of open standards 180 days after the memo. The US DATA Act (2014) requires the creation of a common data schema for the exchange of budget information, and the UK Local Government Transparency Code (DCLG, 2014) is accompanied by strong guidance about the fields that should be used for the disclosure of 14 priority datasets (LGA, 2015). Efforts like the International Aid Transparency Initiative, Open Contracting Data Standard and Budget Data Standard are all working to articulate specific standards for open data publication as part of wider political processes seeking to secure sustained information and data disclosure.

However, whilst advocacy for OGD has focussed on 'big tent' arguments suggesting that the provision of open data brings multiple benefits to a diverse range of stakeholders (Weinstein & Goldstein, 2012), critics have presented the open data movement as a tool for marketisation of public services (Bates, 2012) and as the co-option of otherwise radical transparency and civic-technology activism (Bates, 2013). Practitioners in developing country have questioned the assumptions built into standards promoted as global norms. And current practices around open data have also led to concerns that it will "empower the empowered" (Gurstein, 2011) and thus engender regimes of information injustice (Johnson, 2013). Central to this literature is the argument that the open data movement has been defined mostly by technical considerations, overlooking the political dimensions of the process (Yu & Robinson, 2012; Morozov, 2013) and presuming that the mere provision of data would automatically empower citizens (Gurstein, 2011; McLean, 2011; Donovan, 2012). In particular, Yu and Robinson (2012: 196) denounce the idea that technical criteria, such as the use of open standards in the release of datasets, should be enough to satisfy calls for transparency, writing that: *"An electronic release of the propaganda statements made by North Korea's political leadership, for example, might satisfy all eight of these requirements [Sebastopol principles on Open Government Data], and might not tend to promote any additional transparency or accountability on the part of the notoriously closed and unaccountable regime"*. To these critiques we might also add lessons from science data sharing, to the effect that data standards rarely produce interoperability or interpretability of datasets. Thus any emphasis on machine-readability opens up important conversations about the decisions that are made in constructing data concerning which stakeholders will have their needs prioritised, and how the costs and benefits of adopting standards end up being distributed.

Yet, these critiques noted, the provision of government data under open standards has become a major demand of open data activists. This demand follows a larger history: the Internet protocols were shaped by a discourse on 'openness' of standards. This rhetoric has found a place in a wide variety of movements, asking for software code, hardware, academic publications or governments to be 'opened' to the public by sharing their foundational components (Russell, 2014). However, the demand for 'openness' in standards was not driven only by rhetoric. Open data activists consider that the use of standards facilitates the reuse of data, and gives more specific meaning to demands for machine-readability. But what do these standards and specifications contain? How do they, in practice, ensure or enhance the machine-readability of data? And how does standardised machine-readable data differ from alternative ways data might be shared, shaping in the process who is engaged in open data re-use activity? To address these questions we look in detail at the histories and contemporary implementation of three major standards, used at different levels for opening government data, to understand how they shape both the machine-readability of data, and how they affect wider practices of governmental transparency.

Framework and Methods

Mirroring a common trend in STS research of scholars *"'intervening' while studying science and technology phenomena"* (Karasti et. al., 2016: 4) we enter this field as both practitioners and researchers: involved in initiatives to support open data publication and use practices, whilst also engaged in the scholarly critical study of open data and open government phenomena. Responding to growing discourse on machine-readability and standardisation in the open data field, we sought to identify a series of applications of open data standards in practice, and to apply methods of "infrastructural inversion" to look beyond the surface narratives, and to explore otherwise invisible and ignored work involved in making datasets available as open data.

Three open government data standards are covered by this paper. The first is the CSV (Comma Separated Value) format, which is a general format, used often for tabular or spreadsheet data. The second is specific to the transit field: the GTFS (General Transit Format Specification), offering a schema for transport timetables. The third is the IATI Standard, generated as part of the International Aid Transparency Initiative (IATI), and presenting a schema for detailed disclosure of aid flows. The development of these cases was an iterative process, combining initially independent work from the two authors into a cross-case analysis to draw out key themes and a deeper understanding of the common and divergent labour and impacts implicated in the production of open data according to different standards.

The cases each contribute to understanding different aspects of standardisation. Whilst the broad label 'open data standards' is commonly used to refer to a wide range of different technical artefacts, we note a distinction between standards as *file formats* that enable the exchange of data between systems, without being directly concerned with the semantic contents of the file and standards as *schema*, which are concerned with describing the fields and data structures a file should contain, seeking to enable the exchange of the meaning of the data as well as the data itself. Both formats and schema, at their respective levels, can be used to perform the technical validation of a data file: determining whether it is structured and encoded according to the file formats specification, and whether it meets validation rules set out within the schema. Although specifying the fields and entities a particular kind of dataset should contain can be done in the abstract, in practice, many schemas are directly related to particular file formats. For example, the GTFS schema assumes a CSV file format, and IATI is based upon XML. From an infrastructural perspective, schema then builds upon the "inertia of the installed base" (Star & Ruhleder, 1996: 113) provided by their chosen file formats, incorporating many of the affordances and constraints that those formats provide.

Data collection itself took place between 2013 and 2015, through a series of interviews and participant-observation activities with 'data workers'. We use the term data worker to capture a wide range of roles within government institutions and their associated agencies. For many of our interviewees, their formal job title was not data related, yet their role has come to involve work in managing or directly producing open datasets. For the CSV and GTFS cases, an initial series of interviews were conducted with project managers in charge of executing an open data policies. They were asked with whom they collaborated for the project to identify the second series of interviews, data producers who have released files in an open data portal. These in-depth interviews were conducted in four French local administrations and in an international institution, each of which had launched some form of open data portal. Following an initial round of analysis drawing out the relationship between file formats and data schema, we introduced a further case drawing on participant-observation and interviews with participants involved in the development and implementation of the International Aid Transparency Initiative (IATI), seeking to explore how far findings from the earlier cases applied outside the French context, and with a different base file format from CSV. Throughout our enquiry we have complemented interview data with examination of data artefacts created in the cases, direct observations of project meetings, document analysis, and an examination of the wider literature related to each of the standards we study.

In the analysis that follows, we start our infrastructural inversion by critically examining the history and institutional context of each standard, and how they have been adopted or promoted within the open data and open government field. We then turn to a synthesis of our empirical data to look at how a number of themes emerging from the research play out across each standard.

Three Standards and Their Stories

Comma-Separated Value

In a nutshell, CSV stands for Comma Separated Values and designates a file format for storing numbers and text in plain-text forms. The format itself is agnostic as to what content the files should contain. It consists of plain text with any number of records, separated by line breaks. In each record, there are fields, which are separated by a character, usually a comma or a tab. All CSV files can be opened in a text editor or a browser, but the data will not be represented as a spreadsheet but rather as simple text. As both humans and machines can read these files as easily as text, they are possible to deal in absence of complete documentation. The CSV format predates personal computers: it has been used since 1967 at least by the IBM programming language Fortran, and has been implemented in virtually all spreadsheet software, and in many data management systems. CSV, easy to work with in most programming languages, makes possible to process data through a simple two-dimensional array of values. In particular, CSV is used for exchanging tabular data between programs and systems.

Although open data activists praise it as a robust standard (Pollock, 2013), only recently have efforts been made to formally standardize CSV. In 2005, Yakov Shafranovich, a software engineer, proposed a Request for Comments (RFC) to the Internet Engineering Task Force (IETF), an organization that develops and promotes the use of open standards on the Internet. Although it is now categorized as "Informational" by the IETF, RFC 4180 is generally referenced as the de facto standard for the format of a CSV file. In particular, it specifies that the first line should include a header defining each fields, that any field should be quoted with double quotes and that all rows should contain the same numbers of fields. However, the RFC leaves a number of important issues unspecified, which limits the use of CSV for certain users on two particular aspects. First, valid character sets are not defined, but the RFC suggest using the ASCII characters set, a standard known for favouring English-speaking users, rather than the more comprehensive Unicode (Palme & Pargman, 2009). Second, CSV does not specify how to represent particular kinds of values, such as decimal numbers and dates, even though some countries like France use a comma as decimal separator, and countries vary in the date format they use, risking substantial ambiguity in how data entries such as '11/02/2015', for example, should be interpreted.

Further efforts to standardize CSV are ongoing. In particular the W3C (World Wide Web Consortium) has initiated a working group on CSV based on the observation that " a large percentage of the data published on the Web is tabular data, commonly published as comma separated values (CSV) files" (W3C, 2013). The working group was constituted as part of the W3C advocacy for OGD, promoted in particular by its founder Tim Berners-Lee. It is built out of the fact that the format *"is resisted by some publishers because CSV is a much less rich format that can't express important detail that the publishers want to express, such as annotations, the meaning of identifier codes etc."* (W3C, 2013). The ongoing research of the working group will lead to standard metadata that aims to support the automatic interpretation of CSV files on the web, supporting tools to work around the ambiguities of the format: even if CSV files themselves do not become completely standardized.

Many Open Government Data activists praise CSV for its simplicity and its machine-readability, but they also indicate its limits. Tim Berners-Lee (2010) defined a 5-star grading system in which publishing data in CSV with an open license warrants a 3-star grade. The website 5stardata.info[1] indicates that to publish to CSV format *"you might need converters or plug-ins to export the data from the proprietary format"*. The Open Knowledge Foundation (2013) considers it as the *"most simple possible structured format for data […] remaining readable by both machines and humans"* but highlights it is *"not good for data where structure is not especially tabular"*. More recently, the Open Data Institute (2014), also co-founded by Tim Berners-Lee, has declared that 2014 was the *"year of the CSV"*. It declared that it is *"a basic data format that's widely used and deployed […] but it is also the cause of a lot of pain because of inconsistencies in how it is created: CSVs generated from standard spreadsheets and databases as a matter of course use variable encodings, variable quoting of special characters, and variable line endings."* The organization has published a tool called "CSVLint"[2] which tests if a CSV file is "readable" according to a series of rules, enforcing a set of rules for what a CSV file actually should be, drawing on, but going beyond, the basic RFC specification. The tool is based on the observation that *"CSV looks easy, but it can be hard to make a CSV file that other people can read easily"*.

On a practical basis, the limited standardization of CSV means that opening a file in this format can require the user to understand the complexities of encoding data. When opening a CSV file in most spreadsheet software, a box will often open, asking the user to specify which encoding character set is used in the file, as well as the separator character which delimits fields, and the decimal separator. By default, most spreadsheet software will follow the RFC guidelines but in many situations, users will have to manually change the parameters so that the data is displayed as a regular spreadsheet with properly delimited fields. Users commonly accessing data produced on systems with other localisation settings from their own (e.g. in other countries/language communities) are more likely to encounter such prompts. This box adds frictions for the general public in order to use CSV files. While it allows a level of widespread compatibility across the software tools used by developers, it increases practically the complexity of using this format for the everyday task of viewing data in a spreadsheet, and leads to different experiences depending on the user's locality and language.

General Transit Feed Specification

GTFS (General Transit Feed Specification) provides a schema for public transportation schedules oriented towards facilitating the reuse of transit information by software developers. The need for a common standard was driven by the increasing use by commuters of their phones to plan their trips, as well as the success of online digital maps such as Google Maps and OpenStreetMap. Each GTFS "feed" is composed of a series of CSV files compressed in a single ZIP archive. Each file details one aspect of transit information: transit agency, stops, routes, trips, stop times, calendar, special dates, and information on fares or possible transfers. Not all the files are mandatory but the specification requires specific and detailed fields, which should not vary between published files. In contrary to CSV as a standard format, as a standard schema GTFS specifies much more than just the encoding or the layout of the data: it requires transit agencies to transform their data to common structures and to adopt common terms and categories. While both standards tend to ease interoperability of datasets, GTFS requires transit agencies engage in a process of commensuration, adapting their data against shared metrics (Espeland & Stevens, 1998). This process demands considerable resources, and excludes many aspects of reality rendered by the standard as "incommensurable". For example, whilst it may be possible to describe the type of bus running a route within an arbitrary CSV file, within the GTFS schema such additional non-standard columns would be ruled invalid, and effectively meaningless.

The GTFS standard itself was initially developed by a software engineer from Google, Chris Harrelson, in reply to a request from an IT manager of Trimet, the transit agency for the US city of Portland. Harrelson was working on the current Google Transit project, which included public transit timetables in Google Maps. It appears that, through this collaboration with Trimet, the standard closely resembled the data feeds they already had in use. Had the initial collaboration taken place with another locality, it is possible to imagine that GTFS would have looked quite different. After Portland, more than 400 transit operators have now implemented GTFS and publish their data feed with this standard, making GTFS the most widely used open data standard for exchanging transit data. It is published freely with an open source license, and along with the tools necessary to validate a GTFS feed. Google has dropped its brand from the name of the standard but remains active in its development and continues to extend the number of transit feeds usable in Google Maps.

International Aid Transparency Initiative

The International Aid Transparency Initiative (IATI) was launched in 2008 to develop a common approach for aid donors to share information on their projects, budgets and spending. Following wide ranging consultations with aid donor and recipient countries, the project adopted an open data approach, based on the eXtensible Markup Language (XML) data format in 2011, publishing detailed schemas to set out what information should be shared about aid projects and how that information should be represented. Whilst it was initially developed to meet the needs of government aid donors and recipients, the standard is now used by over 400 organisations, including an increasing number of Non-Governmental Organisations.

Unlike CSV (and GTFS), which use a tabular (two-dimensional) data model, the XML format represents data using a tree structure, where data elements can be nested inside other data elements. It also has a range of in-built mechanisms for validating data, defining value types (e.g. date, number etc.), and standardising how multilingual data should be represented. The XML format was developed by a working group at the W3C between 1996 and 1998, and has since gone through a number of iterations. It is derived from Standardised General Markup Language, which has its roots in the mid-1980s, and itself descends from IBM's Generalised Markup Language (GML), which goes back to the 1960s. The particular innovations of XML include better handling of different character encodings (important for exchange of data containing multiple languages), and new approaches to checking the 'well-formedness' of documents as well as their validity against some defined meta-level schemas (Flynn, 2014).

At the core of IATI is a standard for representing records on individual aid activities. These 'iati-activity' elements can contain project descriptions and classifications, data on project location, budget information, and detailed transaction level reporting of commitments and spending. The standard also allows each activity element to include details of project results, and associated documents. Few elements are made mandatory by the XML schema of the standard, although many are important to have for detailed and forward-looking information on aid. The standard also provides an extensive range of code lists for the classification of activities, some drawn from existing recognised code lists, and others created specifically for, and maintained by, IATI.

In common with many data standards, few aspects of the IATI are completely new. Rather, it was assembled from past precedent, seeking to find a common ground between the existing systems of major aid donors such that it could be at least minimally populated by data already held. The idea of standardised aid information exchange has a long history. Whilst the OECD's Development Assistance Committee Creditor Reporting System (DAC CRS), based on survey data collection of headline statistics from member governments, has been in place since the 1960s, it was in the late 1980s and early 90s that efforts for standardised digital exchange of detailed ongoing project information emerged. The Common Exchange Format for Development Activity Information (CEFDA), a disk-based exchange system, coming before widespread Internet adoption, was the first effort in this direction, although it ultimate saw limited uptake. However, it's field definitions influenced the creation of International Development Markup Language (IDML) in 2001 (Hüsemann, 2001), a format primarily developed to feed data into the Accessible Information on Aid Activities (AiDA) database developed by Development Gateway (initially a World Bank project). IDML and AiDA in turn influenced the development of IATI, both as donors rejected the idea of 'yet-another-database', opting instead for an approach premised on the distributed publication of interoperable data, and as the XML experience of IDML was available to draw upon in building up an IATI standard.

The 'extensible' aspect of XML can also be put to use in IATI, as it allows valid data to embed new fields within the existing structure, declaring alternative 'namespaces' for this data outside of the formal standard. The intent in the IATI case is that this could support de-facto standardisation between small groups of data publishers, without requiring the full process of changing the standard to accommodate use-cases only of concern to a small community of users. However, in practice most extensions to the standard have taken place through the regular revision process, with, for example, more detailed fields for geocoding the location of aid projects recently introduced.

Whilst XML is well suited for exchange of structured data between machines, it can be complex to work with in web applications, and tools exist to help users who are more familiar with tabular data to open and manipulate XML. As a result, IATI has also seen a degree of tool building and secondary standardisation take place, designed to convert the IATI XML data into other formats optimised for different users. A 'data store' has been created which aggregates together known IATI XML files, and then provides various possible CSV rendering of these (each having to choose which elements from the tree-structure of the data to treat as the rows in the file, choosing, for example, between one 'activity' or one 'transaction' per row), and which also offers a JSON (JavaScript Object Notation) format, targeted at web application developers. Each of these alternative formats is in some way 'lossy', containing less information than the XML. Yet, in practice these alternative mediated presentations of the data become the forms that most users are likely to encounter and work with.

Whilst open data standards may often be presented as simple technical artefacts that can be transparency applied to existing datasets, and as a relatively new feature of the open data landscape, these sketches illustrate the long history of even the 'simplest' of standards, and point towards the embedded politics, affordances and limitations of each. We turn then to look at how these standards collide with the work practices of those responsible for making open data available.

The Transformation of Practice: Standards and Data Workers

The use of open standards requires data workers to transform their datasets and to adjust to the standards. This intensive work, led by data producers and open data project managers before opening the data, is rarely measured in advance and is often hidden in the back rooms of open data (Goëta, 2014). As their adoption can require major transformations of pre-existing datasets, standards may increase the complexity of releasing machine-readable and re-usable data. Yet, they can also explicitly or implicitly encode knowledge about how to increase the accessibility of data to a particular community of users.

In order to make a usable CSV dataset, data workers frequently make deep modifications to the original files held by government. The complexity of this transformation is well illustrated by an internal document made by the region Ile-de-France, latter published online as the general guideline of relevance to other organisations releasing open data in CSV format. Entitled "Open data: good practices using Excel"[3], it aims to help data workers publish data in the region's open data portal. The portal policies require data to be published in CSV format, and encourages the geocoding of data entries. Among the recommendations it provides, more than half are directly driven by the specifications of the CSV format. The document asks data workers to fit to the standard, as many aspects of their datasets will simply disappear when changing the format:

> *"One sheet=one dataset": CSV does not support multiple sheets;*
> *"No information should be transmitted by using color—> in CSV format, these data will be suppressed!"*
> *"No merged cells"*
> *"Beware with hidden lines!—> they will display in CSV."*

Besides, the document asks data producers to reorganize the structure of the datasets to fit the RFC specifications of a CSV file which is in use in the region's open data portal:

> *"Column headers on the first lines (=columns titles)";*
> *"No empty cells on columns titles";*
> *"Avoid empty lines or columns";*
> *"Warning with 'orphan' data" designating fields, which are outside of a table and will not display properly in the portal.*

These requirements imply a major transformation of datasets in order to fit it to the CSV standard. The files that officials are being asked to make available under OGD policies were generally not originally produced to be released outside the organization in another format. In the organisations we have studied, it is not the data producers (the subject matter specialists working in the policy areas the data describes) who carry the work of transforming the data to adapt to the CSV format. Instead, open data project managers, whose mission is to actually open the data, take charge of modifying the datasets to fit them into the required formats and standards. In our CSV case, these project managers, originally hired to develop a data portal and foster reuse of the data, have become data managers, directly involved with ensuring the compatibility of specific government datasets with open standards, and interposed between the domain-expert data producers, and the public who access the open data produced. As one explained:

> *"Project manager: When we receive an Excel file, we open it and there are basic stuff such as merged cells, [information in] bold, color…*
> *Interviewer: Do you remove it?*
> *Project manager: Yes, anyway if you want to pass it in CSV, all of a sudden everything disappear and the thing is that for certain files the guys they put color on it although in CSV there is no color. So you have to create other columns.*
> *Interviewer: And how do you do in these cases?*
> *Project manager: Well, you do it manually."*
>
> *(Open data project manager, local authority, France)*

Information erased by the standard has to be rebuilt by the open data project manager who translates this information into a structure that passes the filters of the format or schema.

Creating a GTFS feed also requires intensive work, and a worker to undertake it. Within the organisations we surveyed, transit timetables are contained in numerous information systems and knowledge about their inner working is spread throughout the division of the organisation. Database managers, existing professionals responsible for various data systems in the organisations, needed to undertake complex work explorating the databases to work out how to actually open their transit timetable data. The exploration is made even more challenging when the data has been released following an externally imposed standard. For GTFS, database managers have to dig throughout the organisation to create a proper feed. One interviewee reported how:

> *"For building our first GTFS feed, we released it with the means available because everyone was not ready in the organisation to publish bus data. So it missed, by the time, around 10% of the stops in the feed. It did not have all the schedules and so, with the feedback from developers, it helped us enhance internally the chain to generate a dataset, to enhance the methods upstream, which create the stops in the different software. After around four or five months of work, we actually succeeded in releasing a dataset, which is actually clean and could be used as such by the developers. […]"*
>
> *(Database manager, transit organisation)*

In this case, it took over four months in this case to create a proper GTFS feed, which contained all the bus stops. A long cry from 'raw data now'. Not only did the GTFS standard require a combination of different databases, but the making of the GTFS feed also required organizational work to align data production between the different data producers. Rather than making a pre-existing dataset transparent, the standard demanded the creation of new infrastructural configurations, and resulted in a new dataset, and a new view on government transit activities that had been unavailable before.

The introduction of the standard and production of externally facing data also surfaced weaknesses in existing internal processes. The same database manager reflecting on data errors describes how:

> *"Sometimes it was a mistake by the system so we fixed it. And sometimes it was just a lack of communication between departments. For instance, we figured out that when a stop changed name, there was not always communication from the person who changed the name. […]"*
>
> *(Database manager, transit organisation)*

As a result, short-term process fixes have been documented, and new practices introduced such as asking everyone changing a bus stop name to e-mails across the organisation to get other systems updated. Longer term, however, the external data standard holds a mirror up to the internal infrastructure, and invites consideration of wider changes. As the database manager reported:

> *"Nowadays, to create a GTFS feed, we mix around 6 or 7 databases. Now it works but we still depend that all the databases are up to date. That's why we are thinking on how to build a new information system for buses in which everything is in one database because now this is really complicated."*
>
> *(Database manager, transit organisation)*

This same dynamic of changing organisational activity was also present in the previously discussed local authority CSV cases. One project manager described his ongoing efforts to ask data producers to structure their datasets according to the standard instead of himself actually transforming the files:

> *"Interviewer: Are you going to transform the data every time they are updated?*
> *Project manager: The thing is we try to educate data producers to well structure their files at the beginning. That will avoid us to effectively remake their files every time. […] We are thinking about it because we figured out that we manage the data at the end of the tunnel."*
>
> *(Open data project manager, local authority, France)*

However, both source data and data standards, are often moving targets. Small changes in source data can disrupt well-established processes for data conversation, such as the first time a non-latin alphabet character occurs in a source dataset, or when data in an unexpected character encoding is passed into a database system. Schema such as GTFS and IATI are also updated over time, and conventions around CSV on the web continue to evolve. In the case of one large IATI publisher, technical and policy staffs look monthly at ways to follow the evolutions of the standard and maintain interoperability of the datasets with other data providers:

> *"Our journey has been one of continual improvement, to make sure we keep up with the technical standard as it evolves through 1.01, 1.02, 1.03 and so-on. And also continual improvement by recognising where there are problems with the data and fixing them, but also doing that on a very incremental and month-by-month basis. The fact that we publish monthly means you can do that - you can make a minor improvement that makes a big difference and over time they really accumulate. The other aspect of that would be increasing automation. We're fortunately to start off with a fairly automated process in that our entire IATI dataset gets generated nightly, and we just publish that once a month..."*
>
> *(Technical lead, IATI publishing government department)*

Whilst some aspects of making data commensurate with a standard can be fixed at the interface between internal systems and external data publication, others require changes to the source data itself. For example, the IATI standard invites particular categorisation, description and geolocation of aid activity information, which often requires additional labour from field staff and project officers to supply suitably detailed and structured information and to keep records up to date according to external open data publication schedules, instead of just according to internal management milestones.

Across these cases then we see that aligning with a standard transformed both the data published, and the data publisher: creating new and dynamic organizational structures and practices that did not previously exist. As with metadata standards (Bowker et al., 2010), GTFS, CSV and IATI all produce infrastructural changes in the organisations that adopt them changing technical activities and wider organizational arrangements.

Measuring Performance of Open Government Data Policies

Thus far, we have focussed on standards in terms of interoperability. However, the term standard is often also used in the context of performance standards: checking whether some phenomena measures up against some agreed minimum level of quality, or some criteria for success (Busch, 2011; Bruno & Didier, 2013). The open data standards we have explored have come to be used as means of operationalising assessments of whether OGD initiatives are delivering against principles of machine readability, or against specific transparency goals.

For instance, the Ile-de-France region, in an internal presentation made public to data producers, uses Tim Berners Lee's 5-star scale for Linked Open Data which sets criteria based on the use of open standards for assessing data quality. This internal document considers that the mere use of CSV, instead of the Excel XLS format for example, increases the quality of the dataset without even looking at its content. The same goes for the use of the GTFS format. One interviewee, a database manager in a transit agency explained that with GTFS *"it allows [you] now to have quality data"* (Database manager, transit organisation, France). The UK has gone further in treating a technical assessment of file format as a measure of 'openness', in 2012 introducing to data.gov.uk a feature, which translated the 5-star scale into an algorithm that grades every datasets on the portal. The team in charge of data.gov.uk justified the publicity of the scores, shown on each dataset page as an 'openness rating' and available to explore as an average per publishing agency, as a *"useful driver to improve the data." (Data.gov.uk, 2015)*

Besides being a sign of a quality dataset, open standards are also used to gauge the advancement of the open data program itself. When launching their scoring algorithm, the UK government announced that *"the average openness score for all departments is 52%, based on the percentage of the datasets published by each department and its arms-length bodies that achieve 3 stars""* (Gov.uk, 2012).). Here, the use of an open file format for publishing data is being used as a proxy for the openness of a government department. For the departments which scored poorly, the Cabinet Office announced it will undertake measures to *"improve their performance"* most notably by producing stronger guidance on how to publish data. Yet, such measures don't look inside the dataset to see whether the data is well structured, accurate or meaningful: they simply assess the container.

Along similar lines in December 2011, after the launch of data.gouv.fr, Regards Citoyens, published a blog post in reaction to the new portal. Entitled "Open data: an average grade for a data.gouv.fr under proprietary formats", the article assesses the new open data portal by the standard in use after an examination of the catalogue: *"We were able to find only a dozen of CSV and XML datasets against several hundreds under Microsoft proprietary formats. A serious effort still needs to be accomplished on this matter. According to the norm of the inventor of the web, it is only a small average grade we can grant data.gouv.fr for its launch."* (Regards Citoyens, 2011) This example again shows again the importance being placed upon open formats as a sign of a "good" open data policy. Tim Berners Lee's 5-star rating has become much more than simple guidelines for opening data: publishing data in open formats regardless of the quality or the content of the data is taken to indicate that the open data program is conducted in a good direction.

In the case of the International Aid Transparency Initiative, the IATI standard has also become operationalized as a domain specific measurement tool. The advocacy organisation Publish What You Fund has created an Aid Transparency Index (ATI) every year since 2011. The ATI, originally based on a manual survey of data provision by government aid donors, now uses indicators *"selected using the information types agreed in the International Aid Transparency Initiative (IATI) standard"*, and weights scores on 22 indicators at least 50% lower if the data is available, but not structured using the IATI schema (Publish What You Fund, 2011). The ATI retains considerable components of manual data collection alongside automated assessment of IATI data, yet it points to the way in which standards can influence the way in which assessments of transparency may be carried out.

However, claims about what makes for 'good' open data performance are not neutral claims: they build in broad assumptions about who the primarily users of government data are, and how that data should be used.

Configuring Data Towards Advanced Users

As we have explored at the implementation of data standards in each of our cases, we have found tensions concerning how usable the standardised data is, and by whom. CSV, GTFS and IATI each place emphasis on a particular kind of machine-readability, and their use ultimately aims at reaching a certain type of user who has the capacity to achieve anticipated goals of the open data project. Indeed the main goal of many open data projects is framed in terms of encouraging reuse of published data to create websites, apps and services that in turn will generate economic, social and political value. This requires the data to reach users who have the technical skills to reuse the data, but also the potential capacity to create services without direct funding from public bodies. For the open data project managers we interviewed, there was a common identification of these users as professional developers inspired by the free/open source movement. As such, the use of open standards in the exchange of information was as much about a cultural practice that encourages the development of ecosystems around the published datasets (Russell, 2014), as it was a practical step to make the data easier to work with. When publishing a dataset, an open data project manager in a French city explained she had to make choices on the format she will use:

> *"I put the data in multiple formats to try to find a balanced choice between a very raw format, for example CSV which is something very usable for developers even if it is a bit less for people who just want to see what the data look like."*
>
> *(An open data project manager, French city)*

Here the use of CSV involves a practical choice that orientates the open data policy. The choice of CSV format will increase the frictions for using the datasets by the general public who might be confused by the settings required to open the file. On the other hand, it may appeal to developers who can directly import the machine-readable dataset in the tools they use or create.

The same choice appears in the case of choosing the GTFS format. The complexity of using a GTFS file is far greater than CSV, as the dataset is divided in multiple related files. As a result it is mostly professional developers and transport specialists with the capacity to use data in this format. In a local transit agency, the choice of the format raised some concern that its complexity would reduce the user base to very skilled developers:

> *"We release the data in a format which is really well done, it's a Google norm but it can be very complex to understand. We told ourselves 'the guy who will be able to release an app with that, is going to be really solid'"*
>
> *(A database manager, local transit agency)*

But the choice of this standard was also in many cases driven by demand from developers:

> *"Our problem was that we asked ourselves 'but in format should we publish the data? GTFS?' We did not really know. […] What we did, we went ask the developers but of course we discussed this between technicians and the developers told us 'In our opinion, GTFS is a good format, popular, documented, easy to access, let's start with that'"*
>
> *(An open data project manager, local transit agency)*

This quote indicates that the project managers followed the recommendations of technically skilled developers in order to increase the usage of the data. However, had the project manager been in conversation with other potential data users, they potentially would have had other answers as to what would make the data more usable to them. Open data are thus often being calibrated to the expectations and needs of the users closest to the officials releasing data: these relationships in effect acting against the implicit idea that open data should be configured to be equally open to all.

In the case of IATI, there has been a long journey to bridge the needs of data providers and users. The early publishers of IATI data, and those involved in governing the design of the standard, were large government aid donors, with established ICT departments charged with generating their open data directly out of large Enterprise Resource Planning (ERP) systems or internal project databases. For these users, structured XML was familiar, and allowed flexibility in expressing their data. However, the community of users around aid information is much less technically adept – consisting in analysts who are much more comfortable in using spreadsheets of tabular data than they are coding to work with nested data structures. As an increasing number of Non Governmental Organisations, with much more limited technical capacity, have entered the community of data publishers, there have been substantial efforts towards the creation of tools and services that can work with IATI XML, either providing web interfaces that hide the underlying data formats entirely, or providing conversion tools that convert between IATI XML and CSV flat file formats. Reflecting on supporting publishers and users of IATI data, one member of community noted:

> *"When you start talking about XML and showing people what XML looks like, and [...] [how an] [...] XML file is different to a CSV file, why it's better to use XML rather than having lots of spreadsheets, they tend to start running for the door... But I think there are ways of explaining IATI and also publishing using something like Aidstream [an online publishing platform] where you don't have to even engage with [...] XML."*
>
> *(A member of the IATI support team)*

There is an explicit recognition around IATI of the need for intermediary platforms, which will sit between many users and the data. The same may be said to be true in the case of GTFS, where skilled users configure map-based interfaces or apps to allow others to access transport information. However, whilst GTFS intermediaries tend to be converting data into information, the limited number of people in the community around IATI with in-depth XML skills leads to a layer of intermediaries converting data into data (Davies, 2010), reformatting from a shared structured standard, into a proliferation of non-standardised flattened file formats.

As open government data standards work to configure data towards advanced users, they also introduce other layers of infrastructure and intermediation between citizens and the state. Contrast the direct request to government for information under Right to Information based transparency, in which the citizen is able to demand an account direct from the state, with the mediated access to information presented through OGD. Our point here is ultimately descriptive, not normative: in some cases, the OGD approach may deliver greater effective transparency – but we have to be attentive to the role that the operationalization of machine-readability and through open data standards is playing in shaping who has data, and how they can access it.

Discussion and Conclusions

When looking at the back rooms of open data, the requests that governments 'use CSV', 'publish data as GTFS', or 'adopt the IATI standard for their data' involve much more than a simple operation in which data producers would use the 'save as' menu item and switch the format. Instead, at a variety of levels, standards are substantively shaping not only the production of open data, but are also leading to quiet and localised transformations of bureaucracies. We have seen how standards formats and schema are increasingly becoming indicators of the advancement of open data programs, and adoption of standards as part of open data publication is seen as a crucial part of enacting an open data agenda, realising core principles of making data machine-readable. In response, government officials are engaging in work processes to turn the spreadsheets used on the desktops of their colleagues, and the internal databases from specific departments, into standardised datasets optimised for a particular kind of machine-readability outside of the state: constructing their 'raw' datasets in the process. Mediated via standards, the transparency delivered by OGD reveals one particular rationalisation and representation of the information held inside the state, focussing on machine-mediated transparency, rather than transparency as a relationship between citizen, and account-giving state. The particular affordances of open data formats and standards, with their emphasis on machine-readability, act as a filter on what can or can't be easily expressed as part of OGD transparency.

However, we must also recognise that the histories embedded in the formats and standards being adopted owe as much to politics as they do to technology. Though often appearing as recent creations, open data standards are embedded in much deeper information infrastructures (Star, 1999). Each format and standard we have explored builds on legacy practices going back decades, whether at the level of formatting, as for XML and CSV, or at the level of the categories and classifications built into the standard, as in the case of IATI, and it's reliance on terms derived from OECD political systems, and its data structure defined through a process of political negotiation. These histories are inscribed into each of the datasets created using the standards, although few data publishers or users may be consciously aware of them at the point of publication or use. For GTFS, for example, the fact that it can express timetables, but not public transport performance, for example, is rarely considered when it is selected as *the* format for publishing transport information (Rojas, 2012). Crucially then, and counter to the tone of much open data discourse, the 'openness' of open data does not mean that it is freed from past politics, or from previous generations of technology, which, through their role in defining the information infrastructures from which data is drawn and the standards via which its open incarnation is represented, continue to influence what gets expressed: what is made visible, and what, in effect, disappears when moving data from inside the organisation to the open data domain.

Unlike the negotiated metadata standards for scientific data sharing, shaped within relatively defined communities of practice, the open government data standards we have explored are generally experienced by the data workers of the state as fixed points. Data workers are tasked to implement the standard, and organise their work practices accordingly, but they have limited practical capability to shape the standards around their local needs. Indeed, incorporating user needs in the dataset is often experienced as problematic by data producers and standard-setters (Denis & Pontille, 2013). Divergent user requirements risk disruption to the abstract and idealised application of the standards, and threaten the goal of 'frictionless', globally interoperable data. Data standards can thus come to stand in place of dialogue with the community of potential users of data. Of course, it is not that these standards have no notion of the user. Choices made over the standards in our case studies configure the data (Woolgar, 1991) to advanced users with the skills to open and reuse them. The materiality of machine-readable datasets anticipates the skills and materializes a certain representation of the user (Akrich, 1992). But the standards in use also create multiple data publics (Ruppert, 2012): developers who can reuse the data to create services, advanced users who can open the dataset and do basic analysis and the general public who are expected to benefit from the opening of data only via intermediaries, using what Ruppert (2012) calls "literary technologies" such as visualisations, maps, applications and online services to gain second-hand access to the disclosures made by the state. By making data machine-readable, open data standards, in theory, allow machines to join the cohort of "armchair auditors" (Ruppert, 2012) producing a particular notion of accountability in the transparency agenda of open data policies.

Open government data standards also exist in a context that tightly couples the technical, social, and organisational with the explicitly political: policy commitments have been publicly made against open data principles. Yet, because open data is framed as being published for anyone to re-use, the test of successful publication cannot be any one specific use of the data, but has to instead be a proxy for potential usability of the data. Evaluation of the standardisation and machine-readability of data has increasingly become this proxy within OGD policy making. Yet, setting format standards as the metric for a good open data initiative can leave the content of a dataset entirely out of the picture. Although a focus on schema standards brings in greater consideration of what the data contains, it still stands in for any evaluation of OGD against ultimate goals of creating transparency, in which evaluations would need to address not only the broadcast of data, but also it's effective receipt and re-use.

The account we have offered in this paper provides an initial overview of just three different standards, operating at different levels within the growing landscape of open data. In drawing on empirical work to present a descriptive account of these standards in action at the point of data production, we have sought to contribute an open data component towards the called for development of 'critical data studies' (Kitchin & Lauriault, 2014), unpacking these data standard assemblages, and looking at their materiality within the context of public organisations. This is by no means an argument against adoption of standards: rather, it is an account intended to support constructive and critical approaches to their evaluation and adoption. Further work is needed to trace forward the consequences of these data standards assemblages, and current orientations towards the machine-readability of data, in producing new transparency regimes of the state. In these relatively early days of open government data standardisation, with a new layer of 'open data infrastructure' being built out through the work of policy-makers, technologists and data workers, developing these approaches to bring standards, their stories, and their possible consequences, into view, requires ongoing attention.

References

Akrich M (1992) The De-scription of Technical Objects. In: Bijker W & Law J (eds) *Shaping Technology—Building Society: Studies in Sociotechnical Change*. Cambridge, MA: The MIT Press.

Almklov P G (2008) Standardized Data and Singular Situations. *Social Studies of Science* 38(6): 873-897.

Atz U, Heath T, & Fawcett J (2015) *Benchmarking Open Data Automatically* (No. ODI-TR-2015-000). London: The Open Data Institute.

Bates J (2012) "This is what modern deregulation looks like": co-optation and contestation in the shaping of the UK's Open Government Data Initiative. *Journal of Community Informatics* 8(2).

Bates J (2013) The Domestication of Open Government Data Advocacy in the United Kingdom: A Neo-Gramscian Analysis. *Policy and Internet* 5(1): 118-137.

Berners-Lee T (2009) "The Next Web". TED talk. Available at: http://www.ted.com/talks/tim_berners_lee_on_the_next_web (accessed: 10.11.2016).

Berners-Lee T (2010) Design issues – Linked Data. Available at: http://www.w3.org/DesignIssues/LinkedData.html (accessed: 10.11.2016).

Birchall C (2014) Radical Transparency? *Cultural Studies ↔ Critical Methodologies* 14(1): 77-88.

Borgman C L (2007) *Scholarship in the digital age: Information, infrastructure and the internet*. Cambridge: The MIT Press.

Bowker G C (1994) *Science on the Run: Information management and industrial geophysics at Schlumberger, 1920-1940*. Boston, MA: The MIT Press.

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

Bowker et al. (2010) Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: Hunsinger et al. (eds) *International Handbook of Internet Research*. New York:Springer, 97–117.

Bowker G C & Star S L (2000) *Sorting Things Out: Classification & its Consequences.* Cambridge, MA: The MIT Press.

Busch L (2011) *Standards: recipes for realities*. Cambridge, MA: The MIT Press.

Bruno I & Didier E (2013) *Benchmarking :* L'état sous pression statistique. Paris: Zones.

Cabinet Office (2013) National Information Infrastructure. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/254166/20131029-NII-Narrative-FINAL.pdf (accessed: 10.11.2016).

Chignard S (2013) A brief history of open data. Paris Tech Review. Available at: http://www.paristechreview.com/2013/03/29/brief-history-open-data/ (accessed: 10.11.2016).

Davies T (2012) Emerging Implications of Open and Linked Data for Knowledge Sharing in Development. *Institute of Development Studies Bulletin* 43(5).

Davies T (2013) "There's no such thing as raw data". In: Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13, New York, USA: ACM Press, 75-78.

Davies T (2014) Five critical questions for constructing data standards. Available at: http://www.timdavies.org.uk/2014/02/21/five-critical-questions-for-constructing-data-standards/ (accessed: 10.11.2016).

DATA Act (2014) S.994 - 113th United States Congress (2013-2014): Digital Accountability and Transparency Act.

Data.gov.uk (2015) Five Stars of Openness. Available at: http://guidance.data.gov.uk/five_stars_of_openness.html (accessed: 10.11.2016).

DCLG (2014) Local Government Transparency Code 2014.

Denis J, & Pontille D (2012) Travailleurs de l'écrit, matières de l'information. Revue D'anthropologie Des Connaissances, 6(1):1-20.

Denis J & Pontille D (2013) Parasite Users? Users and Cycling Infrastructures in OpenStreetMap.

Denis J & Pontille D (2014) Parasite Users ? Users and Cycling Infrastructures in OpenStreetMap. In Mongili A & Pellegrino G (eds) *Information Infrastructure(s): Boundaries, Ecologies, Multiplicity*. Cambrige: Cambridge Scholars Publishing, 146-167.

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

Edwards P, Mayernik M S, Batcheller A L, Bowker G C, & Borgman C L (2011) Science friction: Data, metadata, and collaboration. *Social Studies of Science* 41(5): 667-690.

Espeland W & Stevens M (1998) Commensuration as a social process. *Annual Review of Sociology* 24(1998): 313-343.

Fecher B & Friesike S (2014) Open science: One term, five schools of thought. In: Friesike S & Bartling S (eds) *Opening science*. New York: Springer.

Flynn P (2014) The XML FAQ. Available at: http://xml.silmaril.ie/ (accessed: 10.11.2016).

Fumega S & Scrollini F (2011) Access to Information and Open Government Data in Latin America. Available at: https://spaa.newark.rutgers.edu/sites/default/files/files/Transparency_Research_Conference/Papers/Fumega_Silvana.pdf (accessed: 10.11.2016).

G8 (2013) G8 Open Data Charter. Available at: https://www.gov.uk/government/publications/open-data-charter/g8-open-data-charter-and-technical-annex (accessed: 10.11.2016).

Goëta S (2014) The Daily Shaping of State Transparency: Emerging Standards in Open Government Data. Presentation at the ESOCITE/4S Joint Meeting in Buenos Aires, August 20-23 2014.

Gov.uk (2012) New funding to accelerate benefits of open data. Available at: https://www.gov.uk/government/news/new-funding-to-accelerate-benefits-of-open-data (accessed: 10.11.2016).

Gurstein M (2011) Open data: Empowering the empowered or effective data use for everyone? *First Monday* 16(2).

Gurstein M (2012) Two Worlds of Open Government Data: Getting the Lowdown on Public Toilets in Chennai and Other Matters. *Journal of Community Informatics* 8(2).

Heald D (2006) Varieties of Transparency. *Proceedings of the British Academy* 135: 25-43.

Hüsemann S (2001) Information Exchange Platform for Humanitarian Development Projects. Available at: https://www.researchgate.net/publication/237243797_Information_Exchange_Platform_for_Humanitarian_Development_Projects (accessed: 10.11.2016).

Janssen M, Charalabidis Y & Zuiderwijk A (2012) Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management* 29(4): 258-268.

International Open Data Charter (2015) Accessible at: http://opendatacharter.net (accessed: 10.11.2016).

Johnson JA (2013) From open data to information justice. In *Annual conference of the* midwest political science association, Chicago (IL), April 13 2013.

Karasti H, Millerand F, Hine C M, Bowker G C (2016) Guest Editorial: Knowledge infrastructures: Part 1. *Science & Technology Studies* 29(1): 2-12.

Khan S & Foti J (2015) Aligning supply and demand for better governance: open data in the open government partnership. Available at: http://www.opendataresearch.org/dl/symposium2015/odrs2015-paper49.pdf (accessed: 10.11.2016).

Kitchin R & Lauriault T P (2014) Towards critical data studies: Charting and unpacking data assemblages and their work. In: Eckert J, Shears A & Thatcher J (eds) *Geoweb and Big Data*. University of Nebraska Press.

Kitchin R (2014) *The data revolution: Big data, open data, data infrastructures & their consequences*. London: Sage Publications.

Lampland M & Star S L (2008) *Standards and Their Stories: How Quantifying, Classifying, and Formalizing Practices Shape Everyday Life*. Itahaca: Cornell University Press.

Larsson T (1998) How open can a government be? The Swedish experience. In Deckmyn V & Thomson I (eds) *Openness and Transpare*ncy. Maastricht: European Institute of Public Administration.

Lathrop D & Ruma L (2010) *Open government: Collaboration, Transparency, and Participation in Practice*. Cambridge, MA: O'Reilly.

Law J & Mol A (2008) The Actor-Enacted: Cumbrian Sheep in 2001. In: Malafouris L & Knappett C (eds) *Material Agency*. Boston, MA: Springer US, 57-77.

Levien R (1998) The decommoditization of protocols. Available at: http://www.levien.com/free/decommoditizing.html (accessed: 10.11.2016).

Local Government Association (LGA) (2015) Local transparency guidance – publishing data. Available at: http://www.local.gov.uk/practitioners-guides-to-publishing-data (accessed: 10.11.2016).

Millerand F & Bowker G C (2009) Metadata standards: Trajectories and enactment in the life of an ontology. In: Lampland M & Star S L (eds) *Standards and their stories*. Itahaca: Cornell University Press, 149-165.

Morozov E (2013) *To Save Everything, Click Here: To Save Everything, Click Here: The Folly of Technological Solutionism*. Scientific American. New York, USA: Public Affairs.

Murillo MJ (2014) Evaluating the role of online data availability: The case of economic and institutional transparency in sixteen Latin American nations. *International Political Science Review* 36: 42–59.

Obama B (2009) Memorandum on Transparency and Open Government. Available at: http://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment (accessed: 10.11.2016).

Obama B (2013) Executive Order - Making Open and Machine Readable the New Default for Government Information. Available at: https://www.whitehouse.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government-(accessed: 10.11.2016).

Open Data Institute (2014) 2014: The Year of CSV. Available at: https://theodi.org/blog/2014-the-year-of-csv (accessed: 10.11.2016).

Open Knowledge Foundation (OKF) (2011) Beyond Access: Open Government Data & the Right to (Re) use Public Information. Available at:: http://www.access-info.org/documents/Access_Docs/Advancing/Beyond_Access_7_January_2011_web.pdf (accessed: 10.11.2016).

Open Stand (2012) OpenStand: Principles for The Modern Standard Paradigm. Available at: http://open-stand.org/ (accessed: 18.10.2014).

Palme J & Pargman D (2009) ASCII Imperialism. In: Lampland M & Star S L (eds) *Standards and their stories*. Itahaca: Cornell University Press.

Peixoto T (2013) The uncertain relationship between open data and accountability: A response to Yu and Robinson's 'the new ambiguity of open government'. *UCLA Law Review* 60(200): 200-213.

Pollock R (2007) Give Us the Data Raw, and Give it to Us Now. Open Knowledge's Blog. Available at: http://blog.okfn.org/2007/11/07/give-us-the-data-raw-and-give-it-to-us-now/ (accessed: 14.11.2016).

Pollock R(2013) Frictionless Data: making it radically easier to get stuff done with data. Blog post in Open Knowledge International blog. Available at: http://blog.okfn.org/2013/04/24/frictionless-data-making-it-radically-easier-to-get-stuff-done-with-data/#sthash.tsCuAV3h.dpuf (accessed: 10.11.2016).

Publish What You Fund (PWYF) (2011) Aid Transparency Index: Methodology and Data Sources. Available at: http://www.publishwhatyoufund.org/index/ (accessed: 10.11.2016).

Regards Citoyens (2012) OpenData: La moyenne pour un data.gouv.fr sous formats propriétaires. Available at: https://www.regardscitoyens.org/opendata-la-moyenne-pour-un-data-gouv-fr-sous-formats-proprietaires/ (accessed: 10.11.2016).

Robinson D G, Yu H, Zeller W, & Felten E (2009) Government data and the invisible hand. *Yale Journal of Law & Technology* 11(160): 160-175. Available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1138083 (accessed: 10.11.2016).

Rojas F M (2012) Transit Transparency. Available at: http://www.transparencypolicy.net/assets/FINAL_UTC_TransitTransparency_8%2028%202012.pdf (accessed: 10.11.2016).

Ruppert E (2012) Doing the Transparent State: open government data as performance indicators. In: Mugler J & Park S-J (eds) *A World of Indicators: The production of knowledge and justice in an interconnected world*. Cambridge: Cambridge University Press, 51-78.

Russell A (2014) *Open Standards and the Digital Age: History, Ideology, and Networks.* Cambridge: Cambridge University Press.

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

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

Strasser B J (2012) Data-driven sciences: From wonder cabinets to electronic databases. *Studies in History and Philosophy of Biological and Biomedical Sciences* 43(1): 85-7.

Suchman L (1995) Making work visible. *Communications of the ACM* 38(9): 56-64.

Sunlight Foundation (2010) Ten Principles for Opening Up Government Information. Available at: https://sunlightfoundation.com/policy/documents/ten-open-data-principles/ (accessed: 10.11.2016).

Woolgar S (1991) Configuring the User. In: *A Sociology of Monsters: Essays on Power, Technology and Domination*. London: Routledge, 57-103.

W3C (2013) CSV on the Web Working Group Charter. Available at: http://www.w3.org/2013/05/lcsv-charter.html (accessed: 10.11.2016).

W3C (2015) Data on the Web Best Practices: W3C First Public Working Draft 24 February 2015. Available at:: http://www.w3.org/TR/2015/WD-dwbp-20150224/ (accessed: 10.11.2016).

Web Foundation (2015) Open Data Barometer: Second Edition. Available at: http://opendatabarometer.org/assets/downloads/Open%20Data%20Barometer%20-%20Global%20Report%20-%202nd%20Edition%20-%20PRINT.pdf (accessed: 10.11.2016).

Weinstein J & Goldstein J (2012) The Benefits of a Big Tent: Opening Up Government in Developing Countries. *UCLA Law Review Discourse* 38(2012): 38-48.

Yu H & Robinson D G (2012) The New Ambiguity of "Open Government." *UCLA Law Review* 178(2012): 178-208.

Zimmerman A (2007) Not by metadata alone: The use of diverse forms of knowledge to locate data for reuse. *International Journal on Digital Libraries* 7(1-2): 5-16.

Zimmerman A S (2008) New Knowledge from Old Data: The Role of Standards in the Sharing and Reuse of Ecological Data. *Science, Technology & Human Values* 33(5): 631-652.

Zuiderwijk A & Janssen M (2014) Open data policies, their implementation and impact: A framework for comparison. *Government Information Quarterly* 31(1): 17-29.

Zuiderwijk A, Janssen M, Choenni S, Meijer R, & Alibaks R S (2012) Socio-technical impediments of open data. *Electronic Journal of E-Government* 10(2): 156-172. Available at: http://www.ejeg.com/issue/download.html?idArticle=255 (accessed: 10.11.2016).

Notes

1 www.5stardata.info (accessed: 14.11.2016)

2 https://csvlint.io (accessed: 14.11.2016)

3 http://fr.slideshare.net/christophelibertidf/bonnes-pratiquesexcel-cc27juin2013 (accessed: 14.11.2016)

“To Infinity and Beyond!”: Inner Tensions in Global Knowledge Infrastructures Lead to Local and Pro-active ‘Location’ Information

Ayelet Shavit
Department of Interdisciplinary Studies, Tel Hai College, Israel / ashavit@telhai.ac.il

Yael Silver
Department of Interdisciplinary Studies, Tel Hai College, Israel

Abstract

We follow two biodiversity knowledge infrastructures that hold conceptual and practical inner tensions, and we argue that some of these difficulties emerge from overlooking local information and different understandings of the term location. The ambiguity emerges from two basic concepts of space – exogenous and interactionist – that are both necessary yet readily suggest inconsistent practices – global standardization and local flexibility – to organize location records. Researchers in both infrastructures first standardized, digitized and globalized their records, then discovered inner tensions, and finally alternated between globally interoperable and locally flexible records. Our story suggests a broader lesson: since both types of ‘location’ information are necessary; and since vast resources were already invested in globalizing knowledge infrastructures; then investing in local knowledge infrastructures and in alternating between both types of memory practices seems the most rational option, and a good way to resist epistemic injustice afflicting local knowledge in peripheral localities.

Keywords: biodiversity, database, epistemic-injustice

A Brief Introduction to ‘Location’ Uncertainty

As these four special issues have argued, ‘knowledge infrastructure’ is a fundamental emerging concept in Science and Technology Studies encompassing a variety of definitions and case studies, with identifiable common threads across this rich diversity (Karasti et al., 2016). For the purpose of our argument here ‘knowledge infrastructure’ is broadly construed, as resources in a network form (Bowker et al., 2010: 98); and according to Dagiral and Peerbaye (2016: 45): “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".

A number of cases studies – especially those linked to involvement of marginalized groups (Jalbert, 2016; Silver & Shavit, in press) – have considered 'knowledge infrastructures' in this broad sense, as do the two case studies elaborated in this article. One basic and necessary element within nearly any knowledge infrastructure is spatial information. Elsewhere, we have shown (Shavit & Griesemer, 2009, 2011) that 'location' – perhaps the most basic and mundane term in science – holds a basic ambiguity. While employing the same term – 'location' – rigorous records of a biological process use two different concepts of space – *exogenous* and *interactionist* – committed to different epistemic values – generalizable representativeness and comprehensive accuracy – that readily suggest inconsistent modes – global standardization and local diversity and flexibility – to organize location knowledge (Shavit & Griesemer, 2009, 2011). This basic ambiguity is especially relevant for long-term knowledge infrastructures, which have been shown to tackle inconsistent information organization on multiple aspects and dimensions (Karasti & Baker, 2008; Karasti et al., 2002, 2006, 2010). 'Location' ambiguity across long-term studies hinders reliable repeatability of an experiment or survey (Shavit & Ellison, in press) and reproducibility of its results (Ellison et al., 2006).

An *exogenous* concept of space stipulates that organisms' effects on their locality[1] - via their social structure, physiology, metabolism, behaviour and history – and can be safely ignored for successfully modelling and predicting their distribution (Guisan & Thuiller, 2005). An alternative, *interactionist* concept of space, stipulates that these aspects cannot be ignored since organisms and their environments are mutually co-determined.[2] Adopting a certain concept of space signifies a commitment, i.e. an actual expenditure of resources (Gerson, 1998: 25), to certain types of values, of a rational and social character (Longino, 1990, 2004), and entrenched working procedures to coordinate the labour by using technology, i.e. computer-supported cooperative work (CSCW) (Gerson, 2007).

An exogenous space is committed to revealing general distribution patterns, hence valuing *representative and generalizable* data; on the other hand, an interactionist bio-space values a *comprehensive* and *accurate* data-set for a particular location.

An example of an exogenous partition of space is regular quadrats according to randomly-chosen longitudes and latitudes. Human investigators define a system of grid lines – latitudes and longitudes – conventionally located with respect to the Earth's poles, equator, and Greenwich, England as prime meridians, with elevations above or below sea level decided at some arbitrary date. Organisms are located in this framework regardless of their specific behaviour or metabolism and independently of the *existence* of that conventionally imposed description. The organisms do not attend to nor can they exploit, their "lat/long".

An example for an interactionist partition is a polygon of borrows or landscape patches in accord with a gopher's activity or tree's presence. Under this concept, the organisms themselves causally contribute to the organization of the space in which they live. An organism's position will causally depend on, or bear significant relation to, its interaction with its environmental context, i.e., of places modulated or constructed by what the organisms in questions do, what their neighbouring species do, and without regard to the conventions of humans that might study them. In this sense, space becomes also the *product* of the interaction of the organisms and *their* environments.

This article emerges from an on-going involvement with two long term biodiversity case studies. The first case study was mainly conducted by following the MVZ's (Museum of Vertebrate Zoology) surveys across California between 2005 to 2008 yet our research continued until 2013; in the second case study we followed a national survey of "Hamaarag" across Israel from 2004 until 2015. There is no explicitly written method for an involved philosopher of science, but it is an active, interdisciplinary and long-term line of work that builds upon the two basic meanings of 'involvement': care and active engagement. In practice, it means a joint research process of several years, where the philosopher produces a description of a scientific research process based on active participation in routine scientific fieldwork – e.g. collecting spiders, writing trapping results etc. – and asking the scientists working beside her questions that are motivated by their mutual care – not necessarily agreement – on how best to obtain the goals of this particular scientific project and how the project's knowledge infrastructure will best represent it. In addition to fieldwork, the philosopher also sits on Principle Investigator (P.I.) meetings, recording these meetings while intervening with questions that invite the scientist to critically reflect on her description and analysis qua conceptual theoretician. For corroboration, repeated individual interviews with the scientists, each focused on practical understanding of one core concept, were recorded in addition to notes being taken. Each interview lasted one to two hours (typically the latter) and its findings were re-visited throughout the years in order to track conceptual changes. Overall, there were 9 scientists working on the MVZ's project, for which 25 in-depth interviews, 21 P.I. meetings and 6 long fieldtrips were joined. In addition – sometimes in parallel – 9 scientists leading the Hamaarag surveys were similarly followed, via 23 interviews, 10 P.I. meetings and 4 fieldtrips. Knowledge gained from the fieldwork, meetings and interviews initiated a historical examination of how a particular scientific practice and concept came to be. For example, observing a certain method being used in the field, and hearing its rationale of use, triggered a study on its original context of use and disuse. This micro-historical work was done in the MVZ's archive [3], presented online and its hard-copies located in the museum's main gallery, in addition to asking the American or Israeli participating scientists to send all their old emails and meeting minutes regarding that research project and research method. These historical results were later brought back to the scientists for short reflections upon their original thoughts and rationalizations.

Such an involved method may be relevant to scientists, HPS (history and philosophy of Science) and STS (Science and Technology Studies) scholars, as well as to any academic who seeks a more pro-active and interdisciplinary academia. Regarding pro-activity, since biodiversity researchers are often involved in conservation and public engagement, an involved method can easily lend itself to resist epistemic injustice. Epistemic injustice (Fricker, 2007) is a wrong done to someone due to a biased perception of her capacity as a knower. Production of knowledge by academic and laypeople working side by side with mutual recognition gives room for local knowledge that is often silenced, and hence it is one way of resisting epistemic injustice. We also ask whether certain ways to organize the data – e.g. top down versus bottom up – lend themselves more easily to such an involved research.

In the following case studies, the first presents a more exogenous concept of space and a more top down mode of organizing location information while the second – a more interactionist concept of space and a more bottom up mode of organization – yet the scientists in both cases thought well in advance about their knowledge infrastructure. Both found these two concepts of space necessary for an accurate and generalizable location record, yet both first invested most of their informatics resources in fitting their data to a single interoperable data model and later recognized its inherent tension. Both resolved their 'location' uncertainty data in a manner that emphasized the interplay of local workarounds alongside universal interoperability – instead of choosing one or the other – which eventually opened new possibilities for scientific research as well as for resisting epistemic injustice.

Case Study I: The Museum of Vertebrate Zoology

A History of Methodologies in a Natural History Research Museum

The Museum of Vertebrate Zoology (MVZ) was established at the University of California, Berkeley in 1908 by the patron and entrepreneur Annie Alexander and the scientific director Joseph Grinnell (Stein, 2001). Grinnell noticed the rapid demographic and economic changes in California, argued that theses trends unfold a natural experiment in species distribution and evolution (Grinnell, 1917), envisioned his museum as a supplier of facts for describing these changes, and guided by his expert advice on how best to handle them, he described an aim of: "serving as a bureau of information within our general field" (Grinnell, 1935: 2). More specifically, the museum researchers and students were to conduct a series of rigorous descriptions of species and sub-species distributions in the same location over time "with application of the 'laboratory method' out of doors as well as in the Museum" (Grinnell, 1935: 1). The laboratory provided a global method, a "placeless location" (Kohler, 2002), and applying this universal standard to specific places (Kohler, 2012) and to the idiographic narrative style of natural history research had just began. Grinnell was so keen on implementing such new technologies that he defined it as one of the duties of a museum director: "Be alert for improvement of methods in every department" (Grinnell, 1929: 5).

In line with this duty, a huge effort was devoted by Grinnell and the MVZ staff to build standardized, detailed protocols for almost every aspect of work in the museum (down to the kind of ink and paper to use). There was an 8-page written standard for recording observations in a field note journal (Grinnell, 1938) and yet another 5-page protocol specifying the structure of species information on small tags and index cards (Wythe, 1925). This minute procedural decision to distinguish between two techniques to record a species' location – open-ended field notes versus standardized cards – is a crucial point in our story, one we shall return to.

Diligent execution and updating of this distinction enabled the MVZ to function for: "the promotion of wildlife conservation and management on a biologically sound basis of fact and principle," (Grinnell, 1938) and "to establish a centre of authority on this coast" (Grinnell, 1907). The MVZ as a whole functioned in ways aptly described by Latour's (1999) 'centre for calculation', and its specimens as powerful 'boundary objects' (Star & Griesemer, 1989).

In 2001, in preparation for the museum's upcoming centennial the museum vision was re-visited and the idea of a "Grinnell Resurvey" was born (Senior staff, March 28, 2006 and May 1, 2006 interview). Studying this resurvey reveals some of the basic commitments and values entrenched in practice of MVZ researchers and information managers. The MVZ's tradition values rigorous and self-recorded work style. When a trap line is set in the field its specific setting and its method and effort of study are all meticulously recorded in one's field notebook *journal*. There one describes -– and if possible quantifies – properties of the specific locations encountered throughout that day: their landscape, weather, snow level, dominant plants, soil, sampling method and the effort of detection.

In addition, the MVZ held an extensive collection of material objects, i.e. specimens, tagged and stored in cabinets. The tag, sometimes called specimen label, is a small piece of paper attached to a specimen in the field. The tag was the crucial evidence guiding the handling of the specimen later on, upon its arrival at the museum, and its structure and content was specified and standardized (Wythe, 1925).

Once the specimens were brought in from the field, their location as indicated on the tags was entered into the MVZ's collection in the format of index cards and was never supposed to be changed or corrected, "and so, reversely the student [of today] may quickly trace back again from any particular specimen its history, by referring to the card catalogue and field notebook" (Grinnell, 1910: 35). Changing the card wording might break this chain of reference (Gannett & Griesemer, 2004; Latour, 1999). For Grinnell, a specimen without such contextual information is considered "lost. It had, perhaps, better not existed" (Grinnell, 1921:108). To add visual context, thousands of photographs were taken (of habitats, localities and specimens) and hundreds of maps were drawn. All these items were stored in the MVZ archives and all are traceable to each individual specimen stored in the collection, since, Grinnell argued, we never know what type of record will be required in the future (Grinnell, 1910: 34-35).

Grinnell stressed the need to use *both* the narrative, local description in a field notebook journal and the standardized description on a small specimen tag, yet he introduced this distinction only to facilitate the widest utility of collected material. Although standardized information might be sufficient for some taxonomic purposes, the narrative notes might be of broader significance to studies of ecology, evolution and conservation —specimens merely documenting the presence of a given species in an ecological context (Griesemer, 1990).

After Grinnell's sudden death in 1939, surprisingly, little has changed in the Museum's methodology. The primacy of an abstract, context-free point on a universal and standardized grid of longitudes and latitudes, referenced by a number with an unequivocal interpretation, began only when the museum collection was digitized. Throughout the late 1970's the MVZ collection records were entered into a computerized database and by 1998 it was the first collection of modern vertebrates in the world to go online.

One of the forces motivating computerization of records was the passage of several environmental laws in the first half of the 1970s. 'The National Environmental Policy Act' (NEPA), signed on January 1, 1970 by US President Richard Nixon, required that a statement assessing environmental impact (EIS) on species must be filed prior to any major US federal act. The Endangered Species Act (ESA), signed by Nixon on December 28, 1973, likewise created a need for information about species distributions for land developers and business entrepreneurs. Soon thereafter a boom of private companies specializing in assessing environmental impact emerged, and they started arriving at museum collections looking for information. In 1972 the American Society of Mammalogists responded by establishing a committee on Information. That committee, which included an MVZ representative, established a common set of standards for database development across all American collections. In the same year, the NSF founded a new program under which museums could apply for funding of cabinets, fumigation equipment, etc. to maintain their collections.

However, if the MVZ were to continue its role as a "centre of authority," it not only had to store information but also to supply it quickly and efficiently to the public. Luckily, the technology to do just that was already spreading in the life sciences. Mainframe computers became routinely used in the mid 1970s, and the NSF responded by expanding its existing funding program to include information technology. The director of this NSF program, William Sievers, encouraged James Patton of the MVZ and Philip Myers of the University of Michigan to jointly propose a grant to computerize the MVZ's and the University of Michigan's collections and make available a database management system for all other museums. In 1978 they received an NSF grant for retrospective capture of information on the Mammalian collection.

The grant compelled the museums to decide on the types of information to record in the database. Given that the free-text locality information of the field journal would be hard to code in a systematic way, decisions about what information to record in the database entailed trade-offs in future searchability of information about locality and required, in turn, a decision comparing the relative significance of different types of 'location'. Specifically, and practically, the question of what location information to code in the database was whether 'locality' information would be extracted from the field journal, the index card or both? It was then, for the first time, that an implicit *commitment* was made to a *single* concept of space – exogenous from the local landscape and its inhabitants rather than sensitive to it – for recording a species 'location' in the database. From then on, ever-increasing resources were allocated to recording an exogenous concept of location.

One reason for that choice was informatics-based. The information that the database software (TAXIR: Taxonomic Information Retrieval) could query needed to be highly standardized and organized within a single table ("flat file"), in addition to taking as little space as possible, given the processing power and storage limitations of 1970s mainframe computers. The short, standardized descriptive locality recorded on the specimen tag fitted that technical demand nicely, while the intertwined, context-dependent, free-text record in the field journal could only be stored but not searched or queried in a flexible manner. However, the main reason to leave aside the localized field notes did not involve software or hardware. It was the legal and economical burden the EIS's and ESA's put on the protection of species (rather than niches or habitats as Grinnell and others recommended (Grinnell, 1910), hence the NSF's explicit interest – and consequently Patton's and Myer's explicit focus in their proposal – in the *specimen* collection, which – by Grinnell's own stipulation– was available first and foremost from the specimen tag record. For that purpose, the field journal lacked information – such as museum catalogue or accession number – and held vast ecological information that was time consuming to retrieve.

In 1980 the MVZ's database became operable. That is, a person sending a question by mail – e.g. which species were found in Yosemite National Park – could receive a written answer within a few days after his query was entered into the mainframe computer. As a result, queries about a taxon – e.g. genus, species, sub-species – found at a certain point on a map could be answered quickly, while all the environmental, geographical and historical information contained in and distributed among the field journals about that species at that time/space point could not, because it was not machine searchable. De facto, this meant, according to anecdotal comments of current MVZ staff members, that queries about the extensive locality records stored in the field note journals were reduced from that point on.

"Backgrounding" this large source of ecological information did not raise any complaint from most database users concerned with species distribution questions. This implied that an abstract point locality became not only necessary but also *sufficient* for many queries utilizing the museum collection. To be sure, some behavioural ecologists and systematists interested in small-scale questions still routinely read field journal information – typically photocopied and mailed to them by an MVZ curator – yet most queries relied on the database as the primary, and sometimes only, way to describe species location.

In 1997 a new programmer analyst presented a new, relational data model for the collection. This database defined not only multiple search attributes for each specimen record – e.g. its location and name of collector – but also defined relations between these attributes, such as: when, where and who collected that specimen. A relational database allowed flexible queries, and was designed to be complete, i.e. contain records of all specimen tags alongside field journal entries, photos, maps and more. Yet, however ambitious and carefully planned, the database's data model could not interoperate with such open-ended records as the field journals.

In 1998 an online database system was jointly developed with the Alaska museum. "Arctos" is still the largest multi-institutional database of natural history research museums, integrating data from thirteen universities. Now that anyone with internet access could quickly and efficiently query the collection, many more did so, yet only queries about location that assumed a regular grid with standardized meanings for each term, unequivocally (and automatically) assigned to a set of data fields defined by the data model, could be answered by Arctos. The specimen tag records, along with lat/long coordinates, fitted these requirements, while the field journal descriptions did not. As seen, Grinnell's original tags did not mention lat/longs and typically referred to the area around the campground (sometimes even to a whole county). To improve the resolution of these location records in the database, the programmer analyst developed a sophisticated georeferencing algorithm and protocol, which allowed one to assign a GIS map point with a maximum error distance (degree of uncertainty) to each historical locality in the collection (Wieczorek et al., 2004). Finally, a standardized location point seemed to be comparable with current and future location recorded by GPS lat/long methods. It was hoped that whatever uncertainty remained could be reduced by reading the field journals (by now scanned and posted online, but still not searchable), applying auxiliary information to the georeferencing procedure, and thus shrinking the error distance around each point.

Thirty-five natural history museums worldwide record localities via this georeferencing protocol created at the MVZ, attesting to the overwhelming entrenchment of one concept of space as sufficient for recording a location outdoors: an abstract, universally standardized and biologically-exogenous point on a GIS map. Problems arose, however, when someone had to actually *replicate* a visit to the same outdoor location years later by following these lat/long coordinates. This line of fieldwork at first did not turn 'location' into a problem, but only meant more work for those diligent researchers who went the extra mile and interviewed old collectors or read old field notes. What MVZ staff often called "the problem with locality" (Shavit, observation during weekly Resurvey meetings during 2005-2008) did not arise until 'replication' became an institutional problem, i.e. until the "Grinnell Resurvey" project demanded in the spring of 2003 an actual return of various researchers to hundreds of survey sites across California after nearly a century.

From an informatics infrastructure perspective, the late 1990s and early 2000s seemed like the right time for such a move, as new computer technologies became available in the field. For measuring a locality, GPS receivers had become cheap enough to replace the heavier combination of map, compass, and altimeter; and for recording locality information, Palm Pilots and laptops equipped with spreadsheet software increasingly replaced handwritten field journals. The new technologies produced mostly numbers and abbreviations instead of narrative free-text descriptions. These new tools became extensively used in the Grinnell Resurvey project, and consequently the protocols for recording 'locality' in the MVZ were changing in important ways, some of them creating new challenges.

One must record new GPS data fields, e.g. precise longitude and latitude, datum, and device accuracy. This makes sense: without such GPS data-fields, using GIS mapping systems is unreliable, and without GIS maps computers are limited in power to represent and predict species distribution. However, this can also produce a common – and often unnoticed – problem. An MVZ senior naturalist explains: "…if a locality couldn't be located at a [GPS] geographic scale sufficient to be usable by the scale of the GIS layer [representing the spatial distribution of variables such as temperature, precipitation or elevation], then the model derived by the combination of those different data would likely be in error, *the extent of which would not be known*. Georeferenced localities can thus give a false sense of security, unless they are located at a scale appropriate to the other information with which they are associated" (Information manager, interview on September 3, 2008).

To allow interoperability between the georeferenced and the field journal's 'location' descriptions, the journal's information was mined and transformed to a standardized format. Locality information that was sensitive to a given species in a particular ecological and social context was transformed into a set of tables and data fields, each with a standardized meaning and structure. Moreover, location information previously readily integrated with species locality information, such as habitats across the trap line, is now separately mined in order to be incorporated into the database. The increasing prevalence of data standardization in current museum work led most MVZ researchers to record what they regarded as their most important data, in private spreadsheets – the analogue of the old field journal, although they were aware that such data are very likely to become inaccessible after a few years due to obsolete software or lack of metadata.

The net effect of these technology-induced changes in practice and in protocols for data-mining the field journal, actually deepened the gap between these two concepts of space, one exogenous to the research subjects but readily coded in the museum's online information infrastructure, the other sensitive to the subjects and their context, but hard to code and not interoperable between information systems. The result of this data-mining process was several databases on different locations (e.g. Yosemite National Park, Lassen Volcanic National Park, etc.), which, in contrast with implicit expectations, did not successfully link to the main MVZ database. Why? Because history matters: these local databases originated from the notebook narrative culture while the data model of the database originated from structured tags; each type of record was recorded at different stages of the field work, for different objectives, suggesting different data fields for recording locality data, different part/whole relations between data fields, leading to different, non-interoperable formats. Mining information from field journals thus did *not* bring about data interoperability, yet, it *did* further marginalize the concept of space embedded in the journal by rendering researchers even less compelled to invest time and effort in the original field journals.

At this point it may seem the researchers were left with the worst of possible worlds: a globally representative, standardized and mechanically objective (Daston & Galison, 2007) record is heavily used while inaccurate on multiple aspects as mentioned above; whilst a locally comprehensive and judgment-based accurate record is decreasingly accessible as researchers become accustomed to receiving their machine-based answers after 1 minute. Ironically, the harder the MVZ staff tried to apply Grinnell's vision, the faster it seemed in some respects to fade away.

Minding the Gap, Local Workarounds and Universal Interoperability

We have argued so far that examining the history of the MVZ's use of two concepts of space can explain, at least in part, how and why a lack of data and metadata interoperability emerged within the museum's informatics infrastructure. This is one reason why history and sociology can be useful for biologists: minute contingencies, historically entrenched in their routine work, brought about this conceptual gap, and it was the biologists themselves who uncovered this practical and conceptual "problem of locality" through careful study and reflection on their own historical records and documents. In this section, we discuss how their continued attention to institutional history, sociology and conceptual meaning is resolving the problem. We argue that resolution involves minding the gap for the purpose of "bridging" it rather than generally "closing" it, by a practice of "local workable alternation" rather than "universal interoperability".

An institutional response to the locality-interoperability challenge surfaced when the MVZ director, the PI for digitizing MVZ collections, the bioinformatics programmers and the georeferencing manager agreed that the way to "'connect'" the different locality records and make them less vague would *not* be to rewrite them all as various kinds of database records with GPS measurements. Instead of unifying all location descriptions, the MVZ resurvey team decided to return to Grinnell's alternating vision: "These field notes and photographs are filed so to be as readily accessible to the student in the museum as are the specimens themselves" (Grinnell, 1910: 34).

Since 2003, a large portion of the field notes and photographs have been digitized and posted online, yet posting did not make this information readily accessible in the sense one expects of queries to relational databases, because the posted notes were not linked with particular specimens. Since 2007, the GReF (Graphical Referencing Framework) project began to link every specimen in the collection with the journal field note page(s) on which it is described. Trained undergraduate students read the online field notes and whenever they come upon a specimen number, a date or a location, they tag it electronically. Later, a link is made to every place in the database where this number, date or locality is mentioned. The result is not interoperable in the strict sense, because one does not receive a machine-produced answer to one's query. However, a satisfying resolution is indeed achieved since one can click on a link from a single specimen page and reach a page in the journal narrating how it was collected. The researcher can thus quickly work back and forth – alternate – between the two kinds of information, posing structured queries in one and reading free-text descriptions that answer different questions in another.

Since both concepts of space are expressed through differing practices to organize location records – universal standardization and local flexibility – and since both these practices were necessary for re-using Grinnell's and the re-survey information, one can and must alternate between them while juxtaposing their different record types. GReF did not invent workable alternation – Grinnell alternated between tags and notebooks a century earlier – but it did exploit computer technology infrastructure to greatly speed it up and make it widely and freely accessible.

The story we have told here is *not* a part of a global transformation from the theory-driven goal of understanding species distributions to a data-driven goal of practically responding to climate change. Grinnell and his successors shared a vision of a universally useful information infrastructure that was based on their own, centralized contributory expertise (Collins & Evans, 2007) and in that sense, the successors have held true to Grinnell's legacy and initiated the resurvey as a fulfilment of that legacy. But the resurvey participants also brought new perspectives to bear, due in part to the transformations of ecological science, in part to changing technologies – especially the introduction of digital computers, relational databases, global positioning satellites and receivers (GPS), and GIS maps, and in part to changing political interests in climate change and pressures that placed a premium on rapid access to data on species distributions. These scientific, technological, and political changes led to tensions when new methods and protocols were brought to bear ostensibly on the Grinnellian project, which we explored here through the lens of 'location' meanings and records. These changes, however, should not be described as replacement of one set of practices by another, but rather as a more complex articulation of concepts and practices derived from the transformation of ecology, society, technology, and their intertwined infrastructures.

Case Study II: Hamaraag's Landscape Modulator

A Research Project Turning into a Monitoring Institution

In 2001, while the MVZ's senior staff began thinking about an NSF grant that would sustain the Grinnell resurvey, another group of prominent ecologists on the other side of the Atlantic began writing their own ISF (Israel's Science Foundation) proposal on species response to climate changes and to the presence of a landscape modulator (LM). An LM – typically a perennial primary producer – constructs a patch in the landscape that affects abiotic variables (e.g. soil moisture, temperature etc.) around its location and thus may filter the presence of other species from other locations (Shachak et al., 2008).

It began in 1999, with three ecologists, one from a research university and two from Israel's Nature and Park Authority, who agreed on a common theoretical interest: to test the LM model as a way of better explaining and managing biodiversity across different spatial scales. Thinking about the LM model required additional fields of expertise, which added four more researchers from three different academic institutions. They all knew and appreciated each other from years back, with discussions starting more than two years before the actual proposal submission. A first draft was completed and distributed within the group on September 2001, yet it mainly revealed the need for further clarification. Discussions continued and in November 2002 a formal proposal was submitted for the ISF's centre of excellence. Three paths were suggested for testing the LM model – mathematical modelling, experimental manipulation and analysis of observations along the Israeli gradient – yet only the middle section – won funding.[4] Hence, although the researchers originally planned for a national database to facilitate the information emerging from their nationwide experiment, the funding forced them to allocate their own limited private funding for the heavy task of building a group database, which meant that during the planning period and the first year after receiving the ISF grant, the data remained within private excel sheets rather than being shared. On November 2003, when the MVZ researchers set up their red truck and Sherman traps to leave for Yosemite Valley and repeat Grinnell's localities, the ISF grant number 1077/03 became operable and data production and organization began.[5]

Similar to the MVZ's conceptual and practical location-deliberation over how best to repeat their survey, whether to revisit a single trap, a transect line or a nearby habitat, these LMB (Landscape Modulator Biodiversity) researchers discussed whether to re-sample individual traps within a patch, individual patches, patch-types within a plot, a bounded box plot (1000 m^2) or an LTER (Long Term Ecological Research) station (20,000 m^2). By 2005 additional LTER stations joined the group, they re-named the project Hamaarag[6] (in Hebrew 'The Web'), and designed together an official logo with a symbol of Israel-LTER (Long Term Ecological Research). Toward 2007, when both ISF and NSF funding period were nearing their closure, the Grinnell Re-survey team utilized the MVZ's Alexandra foundation funds to maintained their original course, while Hamaarag seemed to be changing its course: some ecologists with a more theoretical stance began to miss P.I. meetings, a few conservation biologists and governmental officials joined, and other founding ecologists changed their titles to 'Board Directors' rather than 'Principal Investigators'.

As a general tendency, the project now became dependent on ministry and private funding.[7] Since January 2007, and especially after the ISF funding ended in October 2008, the field-protocols and data-models became focused on monitoring rather than experimentally testing a model, and the main mission turned out to be providing intensive, rich and reliable data (Karasti et al., 2010) for evidence-based national management rather than a basic theoretical synthesis. To symbolize this gradual change, in 2010 a director who was not a trained biologist was appointed, and in 2013 the new logo lost any mention of the LTER network, the official name slightly changed to Hamaarag, and almost all the data were collected outside the original LMB research stations.

The product of all this monitoring work is, first and foremost, an annual comprehensive report freely downloaded from Hamaarag's website and actively presented to relevant governmental officials. In addition, Hamaarag conducts many other activities to increase accessibility of its results, with an explicit aim to reach policy makers and enforcers at all levels rather than only scientists (Shavit, observation during Hamaarag's meeting on September 4th, 2012 and January 10th, 2013). Hence the dominant printing language is Hebrew for the Hamaarag and English for the LMB. After 12 years, it seems the scientific transformation is complete: from a theoretical and question-driven study to a practical and data-driven information infrastructure for national conservation.

We, the philosophers involved in this study, would agree about the result; we would also agree about the socio-political pressures mentioned above that contributed to this result. All this is relevant, yet we claim *not sufficient*, for telling the story of this project. Tracking the changing structures of the project's information infrastructure will tell us a deeper and more complicated story of memory practices (Bowker, 2005). That is, a knowledge infrastructure, in particular an online database targeting biodiversity in the face of climate change, is not a mere object but a dynamic and context-dependent network of commitments and choices, and the particular structure of organizing its 'location' information can reveal what these researchers are committed to remember and what they choose to forget (Bowker, 2005: chapter 3).

In particular, building an information infrastructure that assumes a single hierarchy of 'part'/'whole' among all LTER stations and a top-down standardization of all the different ways to describe a location, was not consistent with the LM's interactionist concept of space nor the international LTER tradition of diverse and flexible e-data structures (Karasti et al., 2006). This inner tension – to be elaborated below – can explain, at least partly, the project's continuous underuse of its databases and its shift toward a location-based monitoring program. The next section will illustrate our claim that following the structure of the project's memory practice may help to explain this chain of events, and it does not suffice to follow only the politics of funding or data-ownership.

Database Genealogy

This section will illustrate how deserting a top-down information infrastructure and instead enabling a localized bottom-up approach enabled new scientific questions, new working protocols and an opening for an involved citizen-science project within a national monitoring program. In order to support this claim about the relevance of information infrastructure for shaping scientific questions, models and practices, we will now briefly unfold its "infrastructure time" (Karasti et al., 2010).

As already mentioned, during the planning time-period of 2001-2002 the question of organizing the data for analysis of multiple users was raised and discussed, yet none of the P.I.'s had formal experience in information management nor sufficient funding for establishing a long-term, large scale online database from their own private research. In 2005 a bright young student began his third year of studying physics and computer science at the Hebrew University of Jerusalem. During that year, he decided to apply his programming skills, acquired through theoretical training in the university and practical experience in the High-Tech industry. After some time searching he found a place at the department of Evolution, Ecology and Systematics (EES), with a senior theoretical ecologist who was also one of the project's P.I.s (Principle Investigators). In August 2005, the young programmer was hired to design and operate the project's database, and his MSc thesis was supposed to interoperate data from this database to help build a theoretical synthesis that would be relevant for all the P.I.'s involved in the project. He devoted a semester to taking ecological courses and meeting field ecologists, and on January 2005, the programmer presented his initial design, received comments and the general approval of all the P.I.'s present, and began to work on making the database operable.

In May 26, 2005, all P.I.'s received a short email from the senior theoretical ecologist asking them to move onto the next step, that is, send their and their students' data in order for the student programmer to "develop the database *for* the synthesis of the ISF data" (P.I.'s email exchange, May 26, 2005. Our italics). In the long email exchange that followed, concern was raised over the conceptual scale of the student's synthesis and the timetable of its publication. Regarding the scale, some were not sure whether the data currently available could answer in a satisfactory manner such a broad question (P.I.'s email exchange, May 28th, 2005). Others expressed concern over the extent of the original data left unpublished for other members of the group – especially graduate students – if the synthesis were done before their student's manuscripts were sent (P.I.'s email exchange, May 29th, 2005). When the P.I.'s were asked to identify which data sets should be left outside the synthesis, what was left was neither general nor interesting enough for the young programmer and his advisor to work on. When the programmer was asked to distinguish the database from the synthesis, it became clear the former was designed for the latter, hence a strict separation was not practically feasible.

All the senior scientists involved knew each other for years, had mutual appreciation, cared about the project, successfully overcame previous rounds of passionate theoretical debates, repeatedly declared that their disagreements were *not* personal (Email exchange on May 30th, June 2th, and during multiple P.I. meetings) and were committed enough to drive long distances for face-to-face talks on June 17th and 21st, 2005. In short, a resolution seemed certain enough to joke about: "given [assuming?] that we are dealing with a reasonable group of scientists, I am guessing that some compromise is possible" (P.I.'s email exchange on May 30th, 2005).

Yet despite all efforts, a lockdown occurred, partly because the agreed original plan was for only one interoperable database for all LTER stations to test a single unifying theoretical LM model previously discussed for nearly a year. At each LTER station a different P.I. invested much time and effort in producing information to test the same LM model while organizing its location data and metadata differently. Reasons for metadata diversity are themselves diverse: a) the LM species differed among sites (e.g. Common Oak (*Quercus calliprinos*) at the Meron station and Negev Hamada (*Haloxylon articulatum*) at the Avdat station), b) hence constructed patches that looked different and denoted different within-patch-type hierarchies (for example a three layer 'woody'-'periphery'-'open' for Oak trees and a 'woody'-'open' dichotomy for Hamada bushes, Programmer's internal report, December 27th, 2006); c) the fixed plots rendered a lat/long description not necessary[8]; d) the abilities of the project's database manager[9] and finally e) the global LTER network is characterized by a mixed bottom-up knowledge infrastructures (Karasti et al., 2010) and highly diverse data-sets are highly common in eco-informatics (Michener & Jones, 2012).

Given an agreed need for a spatial hierarchy in the database between the sampling units – trap, patch, plot, site, country – but with no agreed mechanism on how to parcel it,[10] and given a single overall synthesis as an agreed common goal but with no agreed temporal mechanism on how to parcel its part/whole relations, a spatio-temporal gap between the P.I.s seemed inevitable. Should the researchers change how their spatial data were organized for the database to be able to automatically aggregate and compare their data? Will it still answer their questions? Should the graduate students donate their data to support the overall synthesis or should the synthesis study await the publication of their results?

Given the effort already invested by the P.I.s and their students in collecting and storing the information in a certain way, it was perhaps rational on their part not to begin investing in alternative suggestions in various possible worlds: two types of synthesis, two interoperable e-infrastructures, or waiting for all the 'partial' or 'small' questions – site or taxon specific – to be answered and published before the overall synthesis would be attended to.

Therefore, although all the P.I.'s wanted to reach an agreement and render the project successful, something had to give, given their goal of a common single model and yet the theoretical and methodological diversification it facilities; and given their goal of a common standard to interoperate their e-data and yet their diverse metadata at each site-location.

Most P.I.s were linked to a LTER station, and thus could continue their work without a uniform organization to their data. Eventually, not enough data were sent to that database and it was never completed. The programming student officially left the project on August 2005, and without an interoperable database to work on, his advisor also became less involved.

It took a while to find a replacement, during which data were curated on a site-by-site basis. On November 2007, a second database manager was hired. This time, no conflict of interests was expected since the database manager had no research interests invested in the database, and data and metadata interoperability was expected since he was a well-experienced informatics person. Indeed, no conflict occurred, yet the new attempt for a single standardized, all-encompassing database, even when detached from any theoretical synthesis aspirations, still did not hold much of the project's data and most of the data it did hold was scientifically underused (Information Manager, interview on December 23, 2008).

Why? One contributing factor might be precisely this detachment. The previous, synthesis-oriented top-bottom database was complex and time-consuming for the biologists to fill, but could potentially test the model they cared about; hence, that effort could be justified. Given that a grand and potentially high profile synthesis was no longer expected to emerge from their data, and given that much of their recorded information could only be standardized via direct communication with their Information Manager and this meant additional work for all those involved, then perhaps it was rational – or at least economic – for many researchers not to invest in changing their data organization or in describing it in detail to their new Information Manager. By 2008, when the ISF grant ended, although several noteworthy publications were indeed produced and the vision of a long-term research was still intact, the completed database was nonetheless left with relatively few data entries (Senior Information Manager, interview on February 28th, 2011). Moreover, each year the database manager received less data from the different LTER sites to standardize and store in the database. In 2010 he also left the project.

For the next three years, there was no database manager and no central database. One may expect that without a unifying information infrastructure to query from, such a national, large-scale and long-term biodiversity project would have surely dissolved. Yet this did not happen. Instead, the project successfully re-invented itself as a national, again long-term, *monitoring* program.

A new, third, database manager was hired for Hamaarag on May 2013. This time, there was no deliberate attempt to use data from the LM research project or its LTER sites. The monitoring data were organized very differently: instead of a single unifying model or a single identical recording protocol to be conducted at all locations, a different, bottom-up scheme was initiated and later coordinated to fit the goal of national monitoring. Hamaarag established teams of experts – theoretical biologists, field naturalists and sometimes policy makers – who specialized in a certain region, habitat or taxon to form a think-tank, defined their specific habitat type, its threats and biodiversity indicators and tailored the monitoring protocol for their habitat and/or geographical region (monitoring director, presentation on March 21st 2013). Hamaarag's scientific committee refined that protocol. Some parts were standardized to fit the protocols of other regions or habitats – e.g. randomly choosing 3 settlements between the variety available as a replicated location of threat – and the resulting location data became standardized and accessible by recording the GPS coordinates of each transect line and exporting that information as a series of KML records to Google Earth.[11]

Some types of interactionist location information – e.g. patch type – were gone, while others remained – e.g. a transect-line located according to an organism's interaction – habitat choice behaviour – with a nearby settlement. Not all ecologists adopt this interactionist perspective,[12] but many do, and the visibility and impact of this project for strengthening an evidence-based national policy for biodiversity conservation is of little doubt. By October 1st, 2014, a fourth information manager arrived, this time immediately following his predecessor and continuing her line of work. Hamaarag, now an official consortium of all the relevant governmental ministries that also cooperated with all major conservation NGO's in the country, began its third year of national monitoring program, secured funds for the next five years, organized its third annual international symposia, published its second "the state of nature report", its first "the state of the sea report", and further deepened its regional approach.

Hamaarag does not seek nor pretend to supply generalizable data that represents a habitat on a national or international scale. Instead, it aims to provide accurate and comprehensive regional data across the country, measuring changes in species richness relative to specific threats and in some cases, offers regional conservation recommendations. As we have seen, this project holds a history of targeting their locations and using bottom up and diverse 'location' descriptions. The targeted location was assumed to be affected by its living inhabitants – human and non-human – and it was the local experts – often different people at different regions – who mostly decided how to characterize a locality and monitor its biodiversity. Given this perspective, it is perhaps less surprising that Hamaarag responded positively to a group of eight policy makers and scientists from a peripheral region who argued for a special monitoring protocol at the northern Hula Valley[13]. Hamaarag's scientists recognized the new questions to emerge from monitoring this small region (which holds 40% of the nation's stream water) and the value of a pro-active municipality. Yet some of the local northern researchers also asked for citizen science information to be considered as part of Hamaarag's knowledge infrastructure,[14] which committed Hamaarag to additional deliberation on the epistemic value of citizen science and of resisting epistemic injustice.

A citizen scientist is a volunteer who collects and/or processes data as part of a scientific inquiry (Bonney et al., 2014; Silvertown, 2009). The citizen science project organized by the local "River-Watch" and the regional "Town Square Academia" added social involvement to the volunteer scientific activity. This local information infrastructure was designed to facilitate a pro-active learning community that would acknowledge and preserve its local heritage. Its data sets were small and diverse yet some environmental protocols were pre-structured to fit the standards of Hamaarag, therefore, enabling a peripheral locality to donate its information to a national infrastructure, and thus receive national recognition of its local expertise and knowledge.

Support for incorporating such local knowledge enables Hamaarag to help resist epistemic injustice (Fricker, 2007), which is, in our case, the injustice inflicted by prejudging the testimony of a resident of a rural periphery, due to her locality, as not really "understanding" her own environment, and therefore not recognizing her as eligible to decide on its future. Between 2009 and 2011 a local NGO named "Nature and Landscape Charms" protested for the people's right for a clean stream running through their city, and in 2012 this local NGO became aligned with another pro-active initiative: "Town Square Academia", who also aimed to recognize the local residence's knowledge about their stream.

The objective of Town-Square Academia is to galvanize an involved and pro-active *regional learning community*. In practice, multiple free courses are conducted outside the campus walls, lead by volunteer experts – academic lecturers together with local people – aimed at conveying existing scientific knowledge as well as documenting and studying local tacit knowledge that is relevant to the community and the researchers. Some courses also develop a group project to continue the learning process and to reach an action-based knowledge within a community of practice (Wenger, 1999). One such course was "A Few Things We Might Not Know About Water" and its group project was the "River-Watch", which still monitors the Jordan River sources, a common resource that physically connects Jews and Arabs, religious and secular, underprivileged and established social groups. Tracking the condition of wildlife and water may empower conservation, build new social and political ties, and suggest an alternative, less-hierarchical and more-involved dialogue between the academia and its locality.

It may seem surprising that Hamaarag, with its nation-wide coverage and uniform standards, would even consider supporting citizen science information. Yet it was considered. An international symposium on citizen science was organized in Jerusalem on February 24th 2014, following a trip to successful European projects. Eventually it was decided that collecting and saving citizen science knowledge would be the responsibility of the SPNI (Society for the Protection of Nature in Israel) rather than Hamaarag. But even after this decision, the head of the monitoring plan at Hamaarag still reserved small funds for the Hula Valley scientists and for the "River Watch" citizen science project. Eventually, a local shortage of determination and funds kept the citizen science knowledge outside the national infrastructure, but why was this option even considered – and with a clear positive spirit – by Hamaarag?

Obviously, we cannot give a definitive answer, yet one possibility is that one's infrastructure can also help entrench a certain theoretical path to be upheld downstream, whether one explicitly agrees with the contingent results of that path or not. In our case, given Hamaarag's rooted use of an interactionist 'location' and its tradition of reliance on local experts holding local information infrastructures; and given the locally oriented request to monitor the Hula Valley alone, with a proactive, human-environmental interactionist concept of space; it became much easier for Hamaarag's governing committee to make place for such an involved location information in their knowledge infrastructure. A similar move would have been much more difficult within a top down, universally and uniformly standardized database. For example, the MVZ's knowledge infrastructure was envisioned and structured, since its establishment in 1907, to be "a centre of authority" in the west coast, i.e. a research institution that spread its own standards of collection and recording rather than absorbing local standards (Shavit & Griesemer, 2011). MVZ researchers systematically relied on local information and opinion, yet a manifestation of this knowledge was never part of the MVZ collection goals and practice, hence it would have been much more difficult to incorporate into its 21th century online database (Shavit & Griesemer, 2011). Given the MVZ's top down approach and Hamaarag's bottom up approach, perhaps it is not so surprising that adding local, pro-active knowledge to the database was positively considered in the latter rather than the former.

Conclusion

This was a story of the various attempts at building and operating a long-term and large-scale knowledge infrastructure, by two influential scientific projects in California and Israel. Both projects thought in advance about how to organize their location information, both have found that two different concepts of space – exogenous and interactionist – are necessary for producing location records that are accurate and generalizable, and both were somewhat surprised to discover inner tensions if both ideals are employed for organizing the same data at the same level.

At first, both invested most of their effort and resources in making their data and metadata even more standardized and globally representative, yet later recognized the inherent tension of recording 'location' only exogenously. Both projects found workarounds to resolve the problem by frequently alternating local and flexible records with global standards – instead of choosing one or the other – and opened new possibilities for scientific research as well as for explication of local memories found – yet not recognized – in the community. Based on these stories we argued that it is justifiable to invest more in the infrastructure to sustain local memories of a locality, and in alternating between the local and global memory practices – both rationally and (sometimes) morally. This argument goes against the mainstream practice of many biologists and database managers, who keep investing ever more increasing funds into streaming and standardizing local data into global databases, as well as keep mentioning its inherent and widespread problems of accuracy (Vanderbilt and Blankman, in press). Following these long term information infrastructures revealed why and how they not only facilitated the preservation of collected data, but also theoretically problematized the foundation of these data and perhaps also directed the future course of its collection and analysis.

Acknowledgments

We deeply thank the MVZ's and Hamaarag's researchers and technicians for their rigorous work and for making this article more accurate. At the MVZ, we especially thank Steve Beissinger, Carla Cisero, Chris Conroy, Karan Klitz, Michelle Koo, Craig Moritz Juan Luis Parra, James Patton and John Perrine. At Hamaarag, we especially thank Nirit Asaf, Naama Berg, David Blankman, Yuchay Carmel, Eli Groner, Avi Pervolotzki Yael Lubin, Ronen Kadmon, Gidi Naaman and Moshe Shachak. Finally, we wish to deeply thank the pro-active people who work with the community and try to transfer that knowledge to the academia: Iris Eshed and Anat Kolumbus. This work could not have been done without the support of the ISF (Israel Science Foundation) grant 960/12.

References

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ & Parrish JK (2014) Citizen Science: Next Steps for Citizen Science. *Science* 343(6178): 1436–1437.

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

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.

Collins H & Evans R (2007) *Rethinking Expertise*. Chicago IL: The University of Chicago Press.

Dagiral É & Peerbaye A (2016) Making Knowledge in Boundary Infrastructures: Inside and Beyond a Database for Rare Diseases. *Science & Technology Studies* 29(2): 44-61.

Daston L & Galison P (2007) *Objectivity*. New York: Zone Books.

Ellison AM, Osterweil LJ, Hadley JL et al. (2006) Analytic Webs Support the Synthesis of Ecological Data Sets. *Ecology* 87(6): 1345–1358.Ellison AM, Barker-Plotkin AA, Foster DR & Orwig DA (2010) Experimentally Testing the Role of Foundation Species in Forests: The Harvard Forest Hemlock Removal Experiment. *Methods in Ecology and Evolution* 1(2): 168-179.

Fricker M (2007) *Epistemic Injustice: Power and the Ethics of Knowing*. Oxford, UK: Oxford Univesity Press.

Gannett L & Griesemer JR (2004) The ABO Blood Groups: Mapping the History and Geography of Genes in Homo Sapiens. In: Gaudillière J-P & Rheinberger H-J (eds) *Classical Genetic Research and its Legacy: The Mapping Cultures of Twentieth-Century Genetics*. Abingdon, UK: Taylor & Francis, 119-172.

Gerson EM (1998) *The American System of Research: Evolutionary Biology 1890-1950*. PhD Dissertation, University of Chicago, IL.

Gerson, EM (2007) Reach, Bracket, and the Limits of Rationalized Coordination: Some Challenges for CSCW. In: Ackerman W.A, Halverson M.S, Erickson C. A, Kellogg Th (eds) *Resources, Co-Evolution and Artifacts: Theory in CSCW*, Dordrecht & New York: Springer-Verlag, 193–220.

Griesemer JR (1990) Modeling in the Museum: On the Role of Remnant Models in the Work of Joseph Grinnell. *Biology and Philosophy* 5(1):3-36.

Grinnell J (1907) *Letter to Alexander*, p. 1, July 7, 1907. MVZ Archive, MVZ Main Gallery, Right Cabinet, File Name: Museum Correspondence - Grinnell.

Grinnell J (1910) The Methods and Uses of a Research Museum. The Popular Science Monthly 77: 163-169.

Grinnell J (1917) Field Tests Concerning Distributional Control. The American Naturalist 51: 115-128.

Grinnell, J. (1921) The Museum Conscience. Museum Work IV: 62-63.

Grinnell J (1929) *Schedule of Curatorial Duties for Staff Members*, p. 5, August 15, 1929. MVZ Archive, MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods - Historical.

Grinnell J (1935) *Analysis of Functions*, p. 2, November 22, 1935. MVZ Archive,MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods – Historical.

Grinnell J (1938) *Suggestions as to Collection and Field Note Taking*, April 20, 1938. MVZ Archive, MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods - Historical.

Guisan A & Thuiller W (2005) Predicting Species Distribution: Offering More than Simple Habitat Models. *Ecology Letters* 8(9): 993-1009.

Jalbert K (2016) Building Knowledge Infrastructures for Empowerment: A Study of Grassroots Water Monitoring Networks in the Marcellus Shale. *Science & Technology Studies* 29(2): 26-43.

Karasti H & Baker KS (2008) Digital Data Practices and the Long Term Ecological Research Program Growing Global. *International Journal of Digital Curation* 3(2): 42-58.

Karasti H, Baker KS & Bowker G (2002) Ecological Storytelling and Collaborative Scientific Activities. *ACM SIGGROUP Bulletin* 23(2): 29-30.

Karasti H, Baker KS & Halkola E (2006) Enriching the Notion of Data Curation in E-science: Data Managing and Information Infrastructuring in the Long Term Ecological Research (LTER) Network. *Computer Supported Cooperative Work* 15(4): 321-358.

Karasti H, Baker KS & Millerand F (2010) Infrastructure Time: Long-term Matters in Collaborative Development. *Computer Supported Cooperative Work* 19(3-4): 377–415.

Karasti H, Millerand F, Christine MH et al. (2016) Knowledge infrastructures: Part I. *Science & Technology Studies* 29(1): 2-12.

Kohler RE (2002) *Landscapes and Labscapes: Exploring the Field-Lab Border in Biology*. Chicago, IL: Chicago University Press.

Kohler RE (2012) Practice and Place in Twentieth-Century Field Biology: A Comment. *Journal of the History of Biology* 45(4): 579-586.

Latour B (1999) *Pandora's Hope: Essays on the Reality of Science Studies*. Cambridge, MA: Harvard University Press.

Levins R & Lewontin R (1985) *The Dialectical Biologist*. Cambridge, MA: Harvard University Press.

Longino HE (1990) *Science as Social Knowledge: Values and Objectivity in Scientific Inquiry*. Princeton, NJ: Princeton University Press.

Longino HE (2004) How Values Can Be Good for Science. In: Machamer P & Wolters G (eds) *Science, Values and Objectivity*. Pittsburgh, PA: University of Pittsburgh Press, 127-142.

Michener WK & Jones MB (2012) Ecoinformatics: Supporting Ecology as a Data-Intensive Science. *Trends in Ecology and Evolution* 27(2): 85-93.

Odling-Smee JF, Laland KN & Feldman MW (2003) *Niche Construction: The Neglected Process in Evolution*. Princeton, NJ: Princeton University Press.

Oyama S (2000) *The Ontogeny of Information: Developmental Systems and Evolution (2nd ed)*. Durham, NC: Duke University Press.

Shachak M, Boeken B, Groner E et al. (2008) Woody Species as Landscape Modulators and their Effect on Biodiversity Patterns. *Bioscience* 58(3): 209-221.

Shavit A & Ellison AM (in press) *Stepping in the Same River Twice: Replication in Biological Research*. New Haven, CT: Yale University Press.

Shavit A & Griesemer JR (2009) There and Back, or, the Problem of Locality in Biodiversity Research. *Philosophy of Science* 76(3): 273-294.

Shavit A & Griesemer JR (2011) Transforming Objects into Data: How Minute Technicalities of Recording Species Location Entrench a Basic Theoretical Challenge for biodiversity. In: Carrier M & Nordmann A (eds) *Science in the Context of Application*. Bielefeld, Germany: Zentrum für interdisziplinäre Forschung, 169-193.

Silver Y & Shavit A (in press) Learning from Success and the Success of Learning: "Town Square Academia." In: Orr Z, Golan D & Rosenfeld Y (eds) *Campus-Community Partnerships in Israel*. Jerusalem: Magness Press [in Hebrew].

Silvertown J (2009) *A New Dawn for Citizen Science*. Trends in Ecology & Evolution 24 (9): 467-471.

Star SL & Griesemer JR (1989) Institutional Ecology, 'Translations,' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907 - 1939. *Social Studies of Science* 19(3): 387-420.

Stein BR (2001) *On Her Own Terms: Annie Montague Alexander and the Rise of Science in the American West*. Berkeley, CA: University of California Press.

Vanderbilt K & Blankman D (in press) Reliable Metadata and the Creation of Trustworthy, Reproducible, and Re-usable Data Sets. In: Shavit A & Ellison AM (eds) *Stepping in the Same River Twice: Replication in Biological Research*. New Haven, CT: Yale University Press.

Wenger E (1999) *Communities of Practice: Learning, meaning, and Identity*. Cambridge, MA: Cambridge University Press.

Wieczorek J, Guo Q & Hijmans R (2004) The Point-Radius Method for Georeferencing Locality Descriptions and Calculating Associated Uncertainty. *International Journal of Geographical Information Science* 18(8): 745-767.

Wythe M (1925) *Suggestion for Handling Specimens Brought Fresh into Museum and Intended for Collections of the Museum of Vertebrate Zoology*, November 13, 1925. MVZ Archive, MVZ Main Gallery, Top Left Cabinet, File Name: Museum Methods – Historical.

Notes

1 'Location' and 'locality' are used interchangeably by the speakers and since this study spans decades and cultures we use an 'actor speech' approach and follow our speakers.

2 Biological models of social-environmental interaction include, for example, niche construction (Odling-Smee et al., 2003), foundational-species (Ellison et al., 2010) and landscape modulator (LM) species (Shachak et al., 2008), while social implications of this approach are explored by Levins and Lewontin's (1985) dialectical perspective and developmental systems theory (Oyama, 1985 [2000]).

3 Additional details on the MVZ's archive: http://mvz.berkeley.edu/History.html

4 We deeply thank the senior researchers for sharing their memories (Interviews on October 18th and 21nd, 2016) and for sending us their email exchanges.

5 The plots were not randomly and independently chosen, as space availability for a plot with the relevant patch types was very limited.

6 'LMB', 'MARAG','HaMARAG' and 'Hamaarag' are all names for more or less the same institution throughout its evolution. It will be referred to henceforth as 'Hamaarag' for the sake of simplicity.

7 From the start, the head biologist of Israel's NPA (Nature and Park Authority), was part of the team. During 2007 the head of the Israeli Academy of Sciences, brought his organization to take unofficial patronage of the project, and one can clearly see the shift towards organizations with a national focus: official letters of support arrived from the minister of Environmental Protection Office, the head for-

ester of KKL-JNF (Keren Kaymet LeIsrael – The Jewish National Fund) and the Heritage Program at the Prime Minister's Office. Given these assurances, the Yad Hanadiv private foundation and later the Ashkol Program, of the Ministry of Science, announced their support.

8 The location of the LTER plots was fixed by fences for a long-term duration, and given the budget constraints and the focus on the plot as the place of repeatable surveys, in all stations except one free aerial photos were used for marking the plot location rather than coordinates from GPS machines.

9 "The information manager usage of the data [is] according to his technical knowledge, for example defining a polygon is more demanding than [defining] a bounding box.**"** Interview with I-LTER database manager. February 28th, 2011.

10 In the programmer's final written report he stresses this point: "The database will *have to* take the patch type hierarchy into account…[but] note that the exact mechanism by which the database will do this is not yet defined!" December 27th, 2006. Our italics.

11 We thank David Blankman for this clarification

12 A dialogue conducted during a 'location workshop' illustrates this point: A statistician: "if we want to do a statistical estimate then in the end you should know the chance to having chosen that location. So the most objective way of doing that is listing every one by one kilometre or five by five kilometre grid cell in the region and then just using a random number generation from Excel or something like that to pick a specific one [location] and say that's your site". The response of Israel's NPA head biologist: "I want to stop you here now, because what you're suggesting is O.K. for a whole [eco-geographical] unit. But there are settlements within this unit, and we see them as the main focus of threat so we need to choose according to them. We cannot do randomly by grid!." October 22nd, 2012.October 22nd, 2012

13 The first meeting was internal to the Upper Galilee people, May 26th 2011, a proposal was submitted to Hamaarag on July 25th 2012 and on July 29th 2014 the cooperation became final.

14 On the very first meeting in May 26th 2011 two of the participants suggested children, students and lecturers as volunteers, and on August 10st 2014 Hamaarag's director sent an email agreeing to embark on a citizen science pilot.

Attachments to Participatory Digital Infrastructures in the Cultural Heritage Sector

Dagny Stuedahl

Oslo and Akershus University College of Applied Sciences, Norway / dagny.stuedahl@hioa.no

Mari Runardotter

Luleå University of Technology, Sweden

Christina Mörtberg

Linneaus University, Sweden

Abstract

This paper explores knowledge infrastructures developed with the aim of opening cultural heritage institutions for public access and involvement. We concentrate on the new modes of knowledge production of professionals and amateur experts involved in the design and use of open archives and wiki communities as a part of transformations towards participatory digital public infrastructures. Ideas of crowdsourcing, policies of open data and engagements in community-based cultural heritage influence participants' visions of future ways of generating, sharing and maintaining their knowledge. The paper identifies how the concept of attachments may help us analytically to understand the dynamics of multiple situated knowledges that are played out when people embrace digital technologies and open-data policies to connect past, present and future orientation of cultural heritage engagements.

Keywords: knowledge infrastructuring, attachments, future orientation

Introduction

The ideas of citizen participation, crowd-sourcing and open data currently receive significant attention from businesses, policymakers, public sector organizations and authorities. The commons movement has drawn interest from the cultural sector, offering new economic, cultural and social models of self-organization and joint activities (Economics and the Common(s), 2013; Hess, 2008; Botero et al., 2012; Björgvinsson, 2014). The emerging interest in open data, crowd-sourcing, and distributed collaboration can be connected to a heightened attention towards social aspects of knowledge sharing, which includes the material, technical and political aspects of how knowledge production is being transformed within a digital cultural heritage paradigm shift (Stuedahl, 2009). It becomes central to understand how ideas of participatory and radical openness are being configured to fit the rapid technology development that is a part of these knowledge infrastructure transformations. We will here focus on how archivists and local historians build on sources and resources from their disciplinary pasts to meet with the present demands and desired futures of open data.

E-government ideals have developed in parallel with the ideas of citizen participation, and open data. The Public Sector Information Directive (PSI Directive) began in 2003 with the goal of actively implementing policies for open data. In 2009 the European Commission stated that, despite progress, there were still barriers hampering the cross-border use of public sector information (COM, 2011b). To advance the market and open up services based on public sector information, the PSI directive has been updated several times (COM, 2011a; COM, 2011b; Directive 2003/98/EC, 2003; Vickery, 2011). The revised PSI Directive includes re-use rules and norms for machine-readable formats and states that governmental agencies should apply open data policies, thereby making data and information available for re-use by businesses, researchers and developers (PSI directive, 2013). As such, open data could constitute a means for infrastructuring an open government. The PSI directive expresses what is to be achieved, but how to actually implement open data and how open data would involve the public in practice, are not touched upon.

Understanding the diverse movements of open data, transparency and citizen participation require a focus on how knowledge infrastructures are enacted in everyday practices where people shape the specific knowledge that keeps institutions together. Knowledge infrastructures are defined as "robust networks of people, artifacts, and institutions that generate, share, and maintain specific knowledge about the human and natural worlds" (Edwards, 2010: 19, Edwards et al 2013). Ever-changing processes of relations, making and practising (Bowker, 1994) prompt a shift to understanding 'infrastructuring' as a verb (Karasti & Syrjänen, 2004; Karasti & Baker, 2004; Star & Bowker, 2002), opening up temporal perspectives for examination. Infrastructuring includes awareness of the procedural (Pipek & Wulf 2009; Bossen & Markussen 2010), and long-term and open-ended processes of continuous co-creation (Ehn, 2008; Björgvinsson et al. 2012; Hillgren et al. 2011; Le Dantec & Di Salvo 2013). These temporal dimensions give a quality of constant 'becoming' to infrastructuring (Karasti & Syrjänen, 2004). As a concept capturing the temporal dimensions of infrastructuring 'the long now' (Edwards et al., 2009; Ribes & Finholt, 2009) has been suggested to point to the long-term dynamics of aligning end goals, motivating contribution and designing for use. This concept focuses on the time scales of actions, and concentrates on the relations and tensions between present demands and a desired future as central for infrastructuring.

This paper argues that this focus on scales of actions between present and desired futures gives an overly limited perspective on what it takes to change knowledge. We argue that knowledge infrastructuring also involves situated knowledge and historical traces and patterns of practices connected to the past, which if not given attention cause simple understanding of tensions. There are currently many expectation and myths (Hellberg & Hedström, 2014) of the positive social impacts of open data. The uptake of open data meanwhile goes slowly, and "the level of knowledge and understanding of open data is currently rather low, and most data producers don't yet see the potential benefits" (Halonen, 2012: 10). There is a resistance to use open data in practice. What if we understand this in relation to the past experience and knowledge that data producers bring to infrastructuring processes? What would a time scale of a longue dureé (Braudel, 1958) including historical time and connections to past knowledge and conceptions of quality bring to knowledge infrastructuring studies?

We apply the concept of attachments to bring attention to how people`s change of practices influence the evolution of knowledge infrastructures over time. Our special interest is in how situated knowledge (Haraway, 1988) may influence knowledge infrastructure processes aimed at openness in the cultural heritage field and how we can understand this from a humanistic perspective on time scales. Attachments point to aspects of knowledge that are not related to content of knowledge, but to value concerns, familiarity and imaginings that may influence infrastructuring (Gomart & Hennion, 1999). In particular, we focus on the role of attachments to existing infrastructuring knowledge practices, standards and values producing tensions with ideas of open, participative and crowdsourced heritage information and resources. Our interest in attachment is based on how this concept may help us reveal how time, especially scales of past, present and future, influence knowledge infrastructuring.

Addressing Attachments in Knowledge Infrastructuring

With the enrolment of attachments we will explore how the complex and multiscale relations and scopes of knowledge infrastructuring are not only about local and situated, global and pervasive, social-material or technological arrangements, but also about scales between past, present and future. 'Attachments' are mechanisms that act as both sources and resources for how people engage, and move between internal motivations and external determinations of knowledge and experience (Gomart & Hennion, 1999; Latour, 1999; Hennion, 2012). The concept of 'attachments' is used in various disciplines: within STS studies, to explore material and social aspects that influence actors in for example public involvement processes (Gomart & Hennion, 1999; Marres, 2007); and within the pragmatist view of Actor-network theory, as a vague and indefinite concept of what holds the social together (Hennion, 2012). It has been suggested as a means to focus on "that which lets/makes happen", giving awareness of passion influencing actions (Greimas & Courtés, 1986, in Gomart & Hennion, 1999). Thus the concept points to entities that are not the content of actions but become apparent in experiences and in people's knowledge practice (Hennion, 2012).

However, the concept has also been used for alternative forms of analysis. In psychology, attachments point to the ties of affection that bind mother and child and endure over time. Another definition is as a distinct unit that adds a function to the thing to which it is connected, in much the same way as the attached document adds a function to an email. In design studies examining the constitution of public, attachments have been used to pay attention to the pragmatic dynamics and fluid social alignments between participants and authorities (Le Dantec & Di Salvo, 2013). Here attachment is applied to capture how sources and resources build 'dependency on' and 'commitment to'.

These affective and pragmatic dynamics of attachments go beyond the articulation work found in earlier infrastructuring studies (i.e. Fujimura, 1987; Strauss, 1988, 1993; Schmidt & Bannon, 1992; Grinter, 1996) focusing on technical and organizational arrangements, and coordination of cooperative work. Also the extended understanding of articulation work as communication and coordination mechanisms, focusing on language and category formation (Baker & Millerand, 2007), may overlook sources that pose their presence in people`s knowledge practices before articulation. Attachments thus provide us with an opportunity to explore entities that are not facts, causalities or reasons but that still have an impact on knowledge infrastructuring. Attachments are both sources and resources for people's agency and knowledge, and direct attention beyond intentions, to entities that do not belong to the vocabulary of action.

Knowledge is attributed to a wide variety of properties and domains, and includes static knowledge of facts, concepts or principles as well as knowledge about valid actions and manipulations within a domain (de Jong & Ferguson-Hessler, 1996). This knowledge-in-use is discussed within both educational and science studies, and serves in both to capture how knowledge is constructed. But, they differ in understanding knowledge as acting on, as opposed to knowledge as actually made (Haraway, 1988). The concept of situated knowledges is used for the knowledge needed to understand situations and to describe how knowledge is partial and situated in local and cultural practices. It is within this understanding of situated knowledges that attachments sit, as entities situated in local and cultural knowledge practices that work as sources and resources for knowledge in the making.

Envisioning the future is relevant for the making of knowledge in the present. Attachments as 'that which lets/makes happen' are also related to how people, in innovation processes, activate knowledge and position themselves in relation to an envisioned future. Thus, transformative knowledge infrastructuring involves re-thinking and re-conceptualizing existing factual knowledge and situated knowledges that have strong relations to the past. While these aspects are sometimes difficult to articulate, some of them might even be tacit: they are attached to how people use their knowledge in transformation processes. Thus, our focus on attachments maps how people experience functions and relations that appear in knowledge infrastructural changes in archives and in cultural heritage communities. This paper explores how attachments are rooted in obligations stemming from the past, which still permeate the present, and how these are related to conceptions of quality in the present and in the future.

Open Cultural Heritage?

The idea of 'open' libraries, archives and museums (LAM) in policy, are highly influenced by the commons movement (Lessig, 2004), where sharing, use and re-use of cultural heritage objects change the focus of these institutions from giving access to cultural heritage into being concerned with how to actively involve the public – or the crowd. Consequently, the concepts of openness and open data are based on the same ideas of public participation as in regulations of digital services delivered by public organizations (Lathrop & Ruma, 2010).

In archives, practices are already distributed and diffuse, and closely related to sets of specialized archival technologies and indexes, which define knowledge in ways that contradict diversity (Bowker, 2005). Meanwhile, archival concepts and practices are transformed radically through technological development – for example, changing conceptions of permanence and stability (Røssaak, 2011). In other words, our contribution to understanding of "the long now" of infrastructuring as implying that long term sustainability requires consideration in the present (Ribes & Finholt, 2009), is the additional consideration of the past. The requirements of transparency, democracy and openness of archives also involve a move towards exposing the contingencies, framing, reflexivity and the politics embedded in these archival technologies (Waterton C, 2010). This infrastructuring aiming at openness and public involvement may cause friction between technological aspects and knowledge practices (Van Passel & Rigole, 2014; Holdgaard & Klastrup, 2014; Perkin, 2010). For example, open cultural heritage initiatives do not only give institutions the options of involving the crowd in, for example, classification, transcription and organization of content (Owens, 2013; Oomen & Arroyo, 2011; Stevens et al. 2010), they also force institutions to re-think archival practices. It becomes necessary to include routines of responding to public users of archival material, and to make space for public contribution and comments on categorizations of archival content. Thus, emerging forms of knowledge practices between archives and the public include re-thinking contextualized relations (Star & Ruhleder, 1994, 1996) and translating current imbrications of formal and informal, ill-structured and well-structured, standardized and 'wild' practice. As such, archives are interesting sites for studying infrastructuring since "information, lived experience and infrastructures" (Star, 2010: 614) are at the centre of their knowledge practices. Focusing on attachments might reveal how conventions of practice emerge and how people negotiate existing dependencies and commitments.

Our approach to infrastructuring for openness, open data, public involvement and crowd participation is concerned with the interpretations that professionals and amateurs make, and how they find ways to realize a new openness within the framework of existing practices of their local institution. We explore infrastructures "in the making" (Bowker et al., 2010) in two different contexts and phases: a startup phase and the implementation phase. Our focus is on how people align envisioned future infrastructural relations with present understanding of what openness involves for their work practice. This affects their translations and engagement, and influences the development of new knowledge-infrastructural relations when technologies change. We present stories from two different case studies drawn from cultural heritage institutions: an archival institution and a

NGO-based local history community. Both groups need to change their knowledge practices when Web 2.0 and 3.0 and social-media-based applications introduce directions and expectations of participation. The two stories exemplify how infrastructuring towards openness and public involvement challenges library, archives and museum (LAM) institutions by introducing new forms of entanglements with the public.

The two stories are about two groups of experts: professional archivists and amateur local historians with shared interests in historical and memory material such as records, photo archives, personal letters, etc. Their knowledge infrastructures and practices differ. Archivists have traditionally concentrated on institutional processes of collecting, preservation and dissemination of historical material; their point of departure is the archive as repository. The local history community works with interpretations and re-presentations of archival material in physical publications. We study how such people imagine potential futures for the norms, practices and routines that they presently share when working to realize a future open and participatory cultural heritage. This informs our understanding that knowledge infrastructuring includes imaginations of the future as well as forms of making and practicing developed from the past. Thus, we ask what attachments people in cultural heritage institutions have to infrastructures of digital heritage participation and open data. How are these attachments related to knowledge and conception of quality connecting to the past? How do these attachments create tension in knowledge infrastructuring processes?

Methodological Approach

We approach these questions with an interdisciplinary collaboration between cultural studies and informatics based on our joint analysis of different research projects in two different Scandinavian countries. The Swedish case is a design project in its start-up phase, focused on translations of the PSI directive, and was especially focused on the way professionals in archives interpret and implement the PSI open data principles in the design of a citizen-centric e-service in archives. This case describes the early stages of knowledge infrastructuring, where people are working hard to translate the directive in relation to existing archival systems. This we define as infrastructuring from above. The Norwegian case of the local history NGO focuses on how local historians convert their writing to a wiki platform. Special attention is paid to how a community re-organize their collaboration, coordination and re-thinking local history genres of writing when using the opportunities of social technologies. We consider the latter as infrastructuring from below involving a knowledge-infrastructuring process where the community has already taken the wiki platform as the premise, and we focus on their handling of the new possibilities and constraints caused by this platform. Thus, the two cases represent two scales of infrastructuring processes and differ in how far the infrastructuring towards openness has been implemented. The Swedish case tells the story of infrastructure in the making from the angle of negotiations in design processes where openness is to be defined and technologies developed to support it. The Norwegian case tells a story about how making involves ongoing configurations in use despite the technological structure in place. The two cases also exemplify two scopes of infrastructuring: the Swedish case is closely directed by policy; the Norwegian case is based on community negotiations related to evolving commitments and dependencies.

The two cases of infrastructuring from above and from below require research methods that capture empirical material to give an understanding of how the infrastructuring processes are experienced differently in different phases. We focus beyond the moments and points of infrastructuring processes where infrastructures become visible or break down (Karasti, 2014). Our aim is to study the formations of infrastructuring, and both studies are based on ethnographic approaches focusing on the infrastructuring as processes of 'becoming', where people build on every-day practices and knowledge "where a need for continuity mandates that new forms emerge through juxtapositions and connections with existing forms" (Karasti, 2014: 2).

The cases have been studied using different methods and introduce different methodological concerns. The Swedish study is based on ethnography of the design process to "facilitate communication and [as] a vehicle for producing information relevant for the design of new products" (Mörtberg et al., 2010: 108). The study is based on participant observation, field diaries and document analysis, and an analytical sensibility towards invisibility of knowledge and experiences, in order to obtain understanding and find ways to integrate and articulate this in technology design (Karasti, 2001, 2003). The study identifies the trajectories of discussion themes and decisions made in the design process, and how these are related to interpretations of directives from policy documents. The study of the Norwegian wiki-community, meanwhile, aims to capture the ongoing infrastructuring in an already-established technical structure. This is based on the principle of not exaggerating the difference between virtual and other settings relevant for the infrastructuring work suggested by virtual ethnography methods (Hine, 2000, 2005). The study focuses on the lived online activity of the community and its connection to offline social spheres (Hine, 2005). Of special concern is how the given technical structure of wikipedia affects or is affected by the transfer from face-to-face community work to online collaboration. This study is based on awareness that online community observations are partial visits into pre-existing processes where isolable and describable locales and cultures are impossible (Hine, 2000), thus including online studies in combination with semi-structured face-to-face interviews. The online ethnography of the wiki-activities concentrated on the discussions connected to writing articles, and followed, from June 2010 till 2012, a thread in the discussion forum connected to one article that had evolved over a long time span. The semi-structured group interview lasted 1.5 hours with 5 staff members of the wiki administration from the institute, including two administrators. The interview was transcribed, and the first publication (Stuedahl, 2011) was shared with the institute and commented upon by the staff members.

In the next subsection we describe two narratives from the case studies, and then we discuss methodological considerations, given that the two studies are based on different disciplinary approaches and scholarly traditions. We then discuss the relevance of the concept of attachments in relation to the case studies and to STS-based studies on public involvement relevant for the activities of infrastructuring for open cultural heritage. We end with a discussion on infrastructuring processes, and how infrastructuring may also contain dimensions of potential future consequences of choices made today, interpreted with knowledge that has cumulated from the past.

Story 1: Infrastructuring the Crowdsourced Archives

The story of the You! Enhance Access to History (YEAH) project can be viewed in terms of the relation between infrastructuring and organizational change (Star & Ruhleder, 1996), and how the issue of creating open cultural heritage data and semantic linking initially was left to the archivist (Runardotter et al., 2011). This project was preceded by the Access to Public Information (APIS) project, 2010–2011, in which the intention was to clarify a research idea and build a network for carrying out a research and development project. The same people formed the project team in both projects: seven people from three National Archives, one from an SME and two researchers from the university, one of whom functioned as project leader, and is also co-author of this paper. In the APIS project there was agreement to collaborate and coordinate material from the LAM sector, and through this to offer citizens re-designed, new and innovative cultural heritage digital services. The APIS project explored the preconditions for creating border-crossing digital services based on archival material by conducting a comprehensive investigation of the area (Runardotter, 2011; Runardotter et al., 2011). Further, it was a collaborative design project with team members from different disciplines. This implies that "different discipline interests [were] brought to the table by each participant" (Baker et al., 2005: 4). Moreover, once agreement on what to accomplish was reached, the design project continued with design activities, such as sketching and modelling and resulted in a demonstrator as a kind of prototype (Bødker et al., 2004). Hence, for participants to reach agreement on a theoretical, abstract level was not problematic.

The YEAH project gained funding in 2011 to address public sector services and their impact on today's society. It set out to focus on public information in archives and citizens' access to it. More specifically, the project aimed to design a prototype of a citizen-centric e-service and explore the role of crowdsourcing methods for augmenting archival e-services with improved access to and usability of archived information. The project team also agreed to design the e-service following three main criteria: it would be citizen-centric, generic and border-crossing.[1] This implied that it would be developed in close collaboration with end-users, building on their needs, requirements and visions, and also through collaboration between different stakeholders (citizens, archives, universities, IT developers, and service providers) and, finally, that the developed e-service should be of interest internationally.

The project members work in public organizations that are distinguished by characteristics such as rational rules and procedures; these organizations have structured hierarchies with formalized decision-making processes and their personnel often advance based on administrative expertise (Parker & Bradley, 2000). These organizations are also subject to political control (in contrast to market control) and therefore cannot be compared with the private sector, since underlying political ideologies influence public organizations' productive activities (Parker & Bradley, 2000). The YEAH project aimed at an e-service in line with intentions stemming from policies around open government and open data, and therefore emphasized the use of ICT in accordance with the European eGovernment Action plan – to help public organizations deliver services to citizens in smart and innovative ways (COM, 2010). As part of knowledge infrastructuring, translations of these policy intentions to the knowledge practices of the National Archives context were needed for the project team to be able to build a prototype that fulfilled the aims, and was in line with the PSI directive.

The YEAH project ran between December 2011 to April 2014. In total the project carried out 32 project meetings, 15 during 2012, 12 during 2013, and 5 during 2013. On 18 January 2012, the project team agreed to aim at "enhancing descriptions of digital objects in existing archival collections by crowdsourcing, in order to improve the description of archival material as well as to improve access to the same".[2] However, public involvement and crowdsourcing was debated among project team members, whose commitment to their professional role and mission became visible in their concern over what this might involve. The discussions continued from January to October 2012 and additional themes, besides crowdsourcing, included whether the archival subject ontology or 'Keywording system' is too complicated for the crowd (that is, any citizen);[3] what if, in using crowdsourcing, the wrong or poor quality information is added (e.g. if wrong person is tagged or if two persons have the same name);[4] how to integrate or link to information created by the crowd with the archival catalogue system[5]. Finally, the decision was taken that the project would develop a framework (demonstrator and guidance) for any memory institution to open their data and link their cultural heritage information to the semantic web.[6]

A central recurrent theme was related to the question of the possible result of citizen-produced material. Issues like "what happens if we let the citizens 'in' and allow them to contribute to our collections" produced a hesitancy towards openness at every monthly meeting from January to June 2012. Another theme was how quality was to be ensured and how quality could be checked when citizens added data, and what could happen if the added information was incorrect. The unpredictability of how crowd-sourced activities and production of content would influence the archival holdings was of great concern to archivists. This gives an insight into how public organizations may be hesitant about new requirements (Kellogg et al., 2006), and highlights how this is connected to the uncertain consequences of crowdsourcing.

As a result, the team decided to review reported experiences of crowdsourced archival material. By June 2012 the project team also reasoned that it would be difficult to make innovative contributions to the LAM communities through crowdsourcing, arguing that "so many crowdsourcing projects have already taken place."[7] This argument shows their main commitment to their own, local organization and community rather than concern for citizen-centric perspectives in themselves. The project team had difficulties figuring out what this project could contribute through crowdsourcing, and also showed uncertainty about whether crowdsourcing was in line with the regulations of open data, as well as how public involvement should happen. Instead, the project team defined their focus in the project as linking open data, and defined this as more interesting and valuable for citizens as an initiative towards openness because of its potential to enhance the experience of cultural heritage data. Consequently, the meeting in October 2012[8] was dedicated to investigating the technical aspects of linked open data, defining the type of archival data the project would work with and how to link it, and how to find a partner with expertise in linking open data. This shift of focus shows how crowdsourcing potentially challenges archival institutions more than Linked Open Data and, since the actual opening of institutional data is still in the hands of the archives holding that data, their feeling of control over future consequences and quality of the content are an important factor in infrastructuring crowdsourced and participatory archives.

Still, it was not until March 2013[9] that the team finally reached a common understanding of what to achieve in the project, and there was agreement on the result of the project: the aim should be to develop 'a demonstrator for any memory institutions to open up their data and link their cultural heritage information to the semantic web.' Decisions were made to focus on genealogists as a representative group of citizens, and to limit the material to be considered. In this way, the project team was able to find a compromise that was acceptable for the project members, as it decreased fear of uncontrolled, or messy, collections in the future. The project ended up with providing a simple methodology to annotate relevant holdings and wrote a methodology handbook on how to create cultural heritage open data and link it to the semantic web.[10] The project ended in April 2014.

The case shows how archivists have several attachments to implementation of openness, and ideas of archives to 'open up' and let data 'out' and citizens 'in'. Several of them are related to the possible everyday consequences of this openness: the accuracy of the data, whether information added by citizens is correct (true), but also to changes in the responsibility for the archival collections and the concern that they are kept in order. In the end, the project turned from a citizen-centric and crowdsourcing involvement initiative towards an inter- and intra-organizational approach that would create less change in the present local-knowledge infrastructure. The eventual form of the policy-driven, and thereby intended, organizational change aiming at openness shows how the archival infrastructure may be challenged when policy requirements for openness meet with interpretations of local- and everyday situated practice.

Attachments to Infrastructuring from Above

The archive story gives an example of how the PSI directive has numerous implications for the cultural heritage sector (COM 2011a; 2011b; Directive 2003/98/EC; Vickery, 2011). In short, the directive is in line with policy goals aiming at an openness that is expected to facilitate democratic processes of the knowledge society, and increase innovation and development of new or improved digital services. This includes expectations of increased productivity and improved effectiveness, efficiency, information quality, interaction mechanisms, better governance tools, and improved government coordination and collaboration (Andersen, 2006; Andersen et al., 2005; Gauld et al., 2009; Grönlund & Ranerup, 2001; Lathrop & Ruma, 2010; Sefyrin & Mörtberg, 2009; Stoltzfus, 2005).

The story tells how knowledge infrastructuring towards crowdsourced archives, following the PSI directive for re-use of public sector information, collides with existing infrastructure relations. Crowdsourcing and public involvement was defined as a less interesting contribution, because it was perceived as less controllable than linked open data. The case underpins the role of attachments for infrastructuring towards openness, and how they are based on professionals' conceptions of control and quality gained from present and local knowledge practices.

In the YEAH project we identify the attachments of hesitancy and unpredictability. These are attachments of a social kind, and can be related to the material attachment of the provenance principle. The principle is fundamental to the careful separation of archives and arrangement according to the original order of the archival collection. The provenance principle is therefore about the importance of knowing where a document was created, by what process, to what end, for whom, when, and how it ended up in the archives (Dollar, 1992). It is important for archives to be able to guarantee the trustworthiness and quality of the information held, as well as making sure that the information is structured and ordered. In the YEAH project, the provenance principle is an attachment from the past that influences knowledge infrastructuring that aims at the future. The reason for holding archives, the preservation of societal memory, relies on this intention of making archives accessible. The intentions of openness in the knowledge infrastructuring had to be scrutinized because of this attachment.

The attachments here were closely related to interpretations of openness in the PSI regulation as related to opening for crowdsourced data, but these attachments did not cause direct actions. Instead they triggered the need for sorting out what openness really means for archives. The translations were multiple; some were hard to grasp, and thus the translations represented an important part of the project narrative. Without common agreement, the project members would not have been able to carry out the design project as intended. It appears that, in order to make progress when infrastructuring, there is a need to make visible and discuss the attachments that people bring along when envisioning the future. Until these are dealt with, it is likely that the people involved will show reluctance, hesitancy and concern. In other words, attachments might be obstacles that hinder development and progress.

Our first story has followed knowledge infrastructuring from above aiming at openness of heritage institutions and in relation to a design project related to the PSI directives of openness. We will now turn to a story of infrastructuring from below from a community-based wiki involving both professional and amateur experts of local history.

Story 2: Infrastructuring in the community based wiki

The lokalhistoriewiki.no was launched by the Norwegian Institute of Local History (NILH) in 2008 after a longer process of trying to implement technology to enhance community activities. The institute is an independent NGO founded in 1955 by the Norwegian Ministry of Culture with the aim of stimulating engagement in local history, and it assembles communities of local history, memory and genealogy (Alsvik, 1993). In 2003 the institute opened a website thought of as a site for collaboration with and between the different communities of amateur and professional local historians connected to the institute. In 2006 they started an online network space for local history projects, aimed at stimulating the sharing of methodological solutions and practical problem-solving of writing local history. However, this initiative ended up being a one-way interaction: the institute serving other institutions, organizations and people. The idea of creating a participatory local history wiki was proposed by members of the Wikipedia community and the institute could see the potential of the wiki format as a solution to connect local historians across different communities and, in addition, assemble the emerging number of local history lexica online. When we approached the project lokalhistoriewiki.no in 2010, the wiki had been up for two years and contained enough online material, articles and forum entries to serve as an object for online ethnographic observations.

Participating in the wiki requires that users register as identified individuals. The wiki-collaboration is supported by four ‘bureaucrats’ from the Institute of Local History constituting the wiki administration in collaboration with 17 ‘administrators’ and 12 ‘vocational supervisors’ recruited from both the Wikipedia community and local communities of amateur historians. The ‘administrators’ have defined roles for sorting out categories and entries, helping new users get started and following up on new publications in the wiki. They have access rights to delete or re-publish pages, they can lock pages, block individual users, edit

Figure 1. In the discussion forum opinions are shared about whether formal or informal categorization makes the wiki most solid, findable and durable over time.

messages and import pages from other wikis. Thus the 'administrators' have a double role as both technical and administrative gatekeepers, and their responsibilities contain both technical and systemic challenges as well as professional evaluations. The 'vocational supervisors' have the role of checking that the articles meet criteria of proficiency, helping users with methodical questions, defining source qualities, sorting out licence questions and answering questions concerning editing or closing pages.

We focus here on how the co-construction of concepts and categories that structure the wiki space evolve. From the discussion forum we have chosen an excerpt from 10–18 June 2010 in which professionals and amateur historians negotiate on establishing a hierarchy of categories for entries and articles on ships, boats and marine vessels. We render the discussion from the thread (see figure 1) in the form of a narrative in order to add relevant contextual information on the trajectory.

This discussion thread was started by one of the professional historians and vocational supervisors, who claimed that categorizing boat types is challenging because the formal categories are built on the registration systems provided by the Directorate of Fisheries, which neglects all the historical boats and vessels in the Norwegian coastal heritage tradition. The participant, anonymized as AK, state the need to make a system of categories that does not need to be reorganized in the future. A group of marine historians from the west coast of Norway has in collaboration with the coastal museum created a structure of categories that builds on a classification system for recording traditional fishing boats in the region. AK provides a hyperlink to their work as a proposal to start the discussion and further argues that this structure embraces information about local occurrences, formal categorization, attribution, name of the boat related to form and function, type of boat, materials used in construction, date, size, volume, name, date of motor, etc. In this way the structure captures both the material and the functional aspects of the boat in question and contains 15 subcategories to be used in the categorizing of boats.

The administrator representing the Wikipedia community, OU, immediately responded to AK´s request and argued that the category system could be built in much simpler ways. OU suggests that marine vessels might be the main category, and then subcategories built on type and function. OU tries to keep the number of categories at a manageable level. After some discussion, OU suggests categorizing the vessels by type. As a response, another member of the community points to the many vessels characterized by their functions (cargo-ships, oil-tanks, service-ships) and that categorization by type only would not cover these. OU proposes categorizing vessels by function, progress and construction.

The next day, OH, a member of the community connected to a museum on the west coast and an active amateur-expert in marine history, suggests a structure of categories that is familiar to people on the coast as well as in maritime communities. OH points out that the index developed by the governmental register is based on well-known acronyms used for more than 100 years, and that these indexes are integrated in the category system at the coastal museum. OH asks whether the category system on the wiki could be developed in correspondence to this well-known category systems – because in writing and storytelling it will be important that the concepts are used in their natural form.

Four days later a member of the wiki administration, IT, starts to build a proposed category three based on type, function, material and construction. This is responded to by Å, who suggests a category denoting the visual form of the boat, observing that categories such as function, material, and construction could provide structure, but might neglect the traditional open boats of Norwegian maritime history. Å ends by pointing out that the wiki should be developed in accordance with a normal thesaurus – and that it is important to clarify this early in the creation of the wiki-structure.

On 16 June, IT asks if the categorizing could start at a more basic level, solely with vessels, and that the category system could be extended gradually when the needs of the wiki community's become more obvious. Collaborator Å answers by asking if it would be fruitful to build the structure of the wiki in relation to concepts used by formal institutions like the Directorate for Cultural Heritage and The Cultural Heritage Act. These use two different categories, separating vessels and boats. The discussion thread ends November 2010, when the administrator IT ask if vessel could be the main category and the sub-categories could be type, use, materials, construction and rig.

Attachments to Infrastructuring from Below

The participants in this wiki-thread are in fact discussing different attachments added to their engagement with the categories of boats and marine vessels to find a 'structure that is common to more than one world to make them recognizable' (Bowker & Star, 1999). They are all well aware that the outcome of this discussion will have a lasting effect on future publications and use of the wiki. It will also determine if this local history wiki becomes interesting for coastal historians and historians of coastal culture. These concerns with the growth of lokalhistoriewiki.no are shared between the wiki administrators, supervisors and contributors and they have a common goal of providing support for multiple forms of knowledge. Meanwhile their attachments to this endeavour are different, and are rooted in different conventions of practice behind their engagement with the lokalhistoriewiki.no project.

The cultural heritage field is characterized by a high level of community initiatives and involvement, long based on participatory knowledge infrastructures where political and social configurations of the community are constantly shifting and hard to grasp (Crook, 2007, 2010). A consensus-based approach to community engagement within cultural heritage has ensured a continued misrecognition of the fact that representations of memories can have the powerful effects of hierarchies, not least in the range of possible stakeholders in the community (Waterton & Watson, 2010). We observe that the local historians involved in lokalhistoriewiki.no see the wiki platforms as an opportunity for a new social and technical platform where issues of authenticity, trust and power may be negotiated openly (Waterton E, 2010), and power-laden hierarchies neutralized. The attachment of the local historians to openness therefore involves keeping the wiki inclusive for multiple historical perspectives, and to build a platform for the community to grow. While their attachments open up to the community of local history, they collide with other attachments to wiki platforms and practices.

The content creation of wiki-spaces is always an incomplete and continuing process relying on constructive participation by the community (Bruns, 2008). From our story we see how this incomplete and continuing process is based on negotiation between local historians and the technical preferences of the wiki administrator that seems to end up with a question of priority of attachments and authority to find solution. The main attachment for wiki administrators is a concern to keep the number of categories below a complexity threshold. The administrator's concerns are directed by basic principles of hierarchical structuring of wiki content, as well as conventions, guidelines and templates for ordering knowledge in wikis to secure simplicity and searchability (van der Velden, 2013). The attachments of the wiki-administrator are related to the organization of the body of knowledge and how the articles relate to each other in the wiki structure, and categorizing structures as a vehicle for connecting the community. Thus, the orderliness of the wiki is a central attachment for the wiki administrator, and he seems less concerned about the quality of the wiki as a matter of giving room for multiple knowledges. The attachments of wiki administrators, therefore, are about commitment to the hierarchical structure of the wiki and the conventions provided by the Wikipedia community.

This is at odds with the concerns that contributors and supervisors from the local historian community share. Their attachment to the wiki is that the category system should fit with and connect the diverging, multiple disciplinary communities that engage with boats and marine vessels. Their rationale for this is to keep a high disciplinary level of categorizing in the wiki structure, and to align well with the ones active among local as well as marine historians. The disciplinary concerns introduce a complexity related to the historical development of boats and vessels, as well as concerns of formal and informal categories. While the open structure of the wiki platform in theory provides the technological means for negotiations across communities, making visible multiple contributions to local history, giving access to participation in discussions from multiple viewpoints, they are still dependent on aligning these with the technical structure of the wiki. Solving the dilemmas of establishing a category structure that is on one hand both technically simple and easy to use for all and on the other hand precise enough to make the wiki suitable for professional knowledge building, it seems priority is given to a technical solution. The disciplinary and professional concerns of the local historians have to align with the ones of the wiki administration.

The story tells how infrastructuring and wiki policies involve negotiations between these diverse attachments. Attachments to developing a neat and simple hierarchical structure that may be used by many collide with attachments to making room for the scope of cultural heritage knowledge and the multiple understandings of boats and marine vessels that connect coastal historians, marine historians and experts of traditional coastal knowledge.

Discussion

Our stories describe two different knowledge infrastructuring processes towards participatory cultural heritage. The Swedish project intended to create citizen-centric digital services on archival material in a public institution in line with the PSI directive. The lokalhistoriewiki.no project in Norway involves infrastructuring in a non-governmental setting heading towards community-based development of local history writing. While both stories relate to processes of infrastructuring aimed at fitting with contemporary ideas and directives of crowdsourced, participatory and open cultural heritage served by technologies, they differ in their approach from above and below. They tell also of different phases of knowledge infrastructuring. The first case describes the start-up phase of deciding how to design future e-services. The second case focuses on knowledge infrastructuring when the work is ongoing. The stories also differ in relation to the materiality of the technology involved: the archivists are relating their design-oriented endeavours to existing technical archival infrastructures, while the local historians have made a decision to use the ready-made technical platform of wiki and configure this to their needs. Thus, material technology plays different roles in the knowledge infrastructuring in the two stories.

But there are also similarities; both stories involve people with great engagement with and passion for cultural heritage: the present, past, and future of cultural heritage understanding and its relevance to society. The two stories also make visible the inherent conflicts that arise when existing conceptions of openness and democratic cultural heritage institutions are contested by technological developments that introduce more radical forms of participation. Archives have always been 'open' in that it is possible to visit and use the physical archival holdings. This openness is taken for granted as part of archives' *raison d'être*. But openness means something slightly different in analogue contexts of managing archival holdings compared to openness in digital infrastructures and practices which brings multiple new dimensions and meanings. The local-historian community is encountering a similar change. Their work has been related to writing paper-based books and journals distributed through local historical societies or by The Norwegian Institute of Local history. The access to local history source material as well as the open distribution of local historical production is imbricated with the collaborative and value laden social space of the community. These social aspects of local historical material bring the *raison d'être* for local historians, and maybe even more for the amateur local history expert. Thus, it is crucial for both cases, as for every memory institution, to address, problematize and sort out the complexity of openness described in policy documents in relation to local practices and their social and material scales.

We have focused on how attachments to openness and public involvement influence infrastructuring in a formal institutional context and in an informal context of an NGO. Our focus on attachments has been related to aspects of knowledge that are both sources and resources brought under scrutiny in knowledge infrastructuring processes. Our stories show that both the open archive and the wiki are objects of study that 'arrive', rather than objects that are performed (Gomart & Hennion, 1999). The Swedish archivists started with designing crowdsourcing possibilities and arrived at focusing on linked open data. The radical understanding of openness came with too many attachments to the unpredictability of crowdsourcing. These were related to control of archival content and quality in accordance with archival provenance principles and concerns with orderliness. These are concerns developed over time, and situated in daily practices that guarantee the trustworthiness and quality of archival information. The story tells how their situated practice of provenance principle collided with principles of crowdsourcing, and how linked data was found as a relevant solution for archival openness. The reasons for these attachments were hard to grasp, and it was hard to understand the scale and scope of consequences the archivists saw of crowdsourcing. There are many understandings and misunderstandings of crowdsourcing in the cultural heritage sector. These are often limited to instrumental understandings of crowdsourcing as an instrument enabling better delivery of content to end users. A re-conceptualisation could for example be in understanding crowdsourcing as an opportunity to actually engage users in public memory in meaningful ways, and as the "fundamental reason that these digital collections exist in the first place" (Owens, 2013: 128).

The story of the Norwegian local historians collaborating with wiki administrators' depart from a context where crowdsourcing is already happening, and is perhaps an example of how practices of the open archive could have become. We see how this open crowdsourcing is a process of constant arriving, of constant negotiations of the categorization structure between technical and social considerations. The attachments of the wiki administrator are closely related to issues reminiscent of the archivists': orderliness, control and quality. In wiki terms, this is about technically simple structure and categorization. These attachments that the wiki administrator connects to growing the wiki conflict with the social and political attachments of the local historians and the way they envision the wiki project bringing together multiple local history communities. By focusing on the diverging attachments at play, we see how one knowledge-infrastructuring process, aligning all actors to one shared and simple structure, collides with the other knowledge-infrastructuring process, making a structure that enrols multiple actors.

The analysis shows how applying attachments as an analytical tool requires a focus on the events in infrastructuring processes, rather than on the relation between agent and structure. Attachments have helped us to pay attention to infrastructuring as a process-in-the-making. In addition, the concept has made us capable of moving beyond intentions and causations to search for entities that do not belong to the vocabulary of action. It is this move that allows us to understand the subtle distinctions between the archivists and the wiki-historians. In both stories we see how attachments towards openness point to different conceptions of orderliness, quality and control. The knowledge infrastructuring in the archival case illustrates how attachments are closely related to situated knowledges (Haraway, 1988) that do not come to terms with radical principles of openness. The knowledge infrastructuring in the wiki case builds a priori on a version of radical openness but illustrates how different sets of attachments cause collision between wiki administrators and local historians. Consequently the two stories of attachments show how situated knowledges sometimes come up short in innovative infrastructuring processes, and may cause good arguments for resisting change or new opportunities for sharing and connecting information and resources.

We have shown how attachments give us access to the negotiations that run the risk to be invisible and not articulated in the policy implementation process. The challenge of tensions, opposition, resistance, work-arounds, and non-adoption of new knowledge infrastructures is not a new narrative in knowledge infrastructuring studies. However, we argue, that the concept of attachment has helped us to pay attention to how relations between past, present and future may sometimes bring contradiction between situated knowledges, the past and the envisioned future. For example policies such as the PSI directive are first and foremost focused on anticipated outcomes: the 'wants' and 'needs' they are expected to fulfil. But we argue that there is no straightforward way ahead.

Existing knowledge practices are sources for understanding the future and our story shows that attachment to contemporary practices and understandings of future scenarios are linked to the history of professional practices. By focusing on attachments we have identified that openness is understood in two ways: first, as a technical concern with 'Open data' or 'Open Software' which indicates that data should be freely accessible to everybody for any purpose, and without restrictions of control besides European and national privacy legislation, such as copyright or patents. This is a technical-administrative understanding of openness. Second, in cultural heritage institutions, openness also relates to 'how' the institution works, which may include the way the institution motivates and aligns multiple goals in collaborations with communities. This is a social understanding of openness. Both understandings are challenged by contemporary ideas of open data, crowdsourcing and emerging conceptions of future democratic institutions. The notion of attachment has given us a tool to identify entities, sources and resources of how these challenges activate situated knowledges that may go beyond articulation work, and include future orientations that are based in historical constituencies of practice but that have still not a developed language.

We started out by asking what attachment people in cultural heritage institutions have to infrastructures of digital heritage participation and open data, and how these are related to existing knowledge practices. We have found that the open data initiative challenges the cultural heritage sector, which has mainly concentrated on preservation, meaning 'taking care of', 'attending to' and 'safeguarding' archival collections. Consequently, openness or closeness seems to relate to knowledge practices of inclusion and exclusion of data, categories and the subsequent knowledge exchange with citizens. In innovations towards openness, these knowledge infrastructures are put under scrutiny, and various risks are attached to social dimensions of public involvement implicated in openness. We have seen how the existing knowledge structures and conventions of practice in archives have no room for attachments that require spaces of contestation (Barry, 2001) and do as a result collapse and bring ambivalence into innovative knowledge infrastructuring.

References

Alsvik O (1993) *Local history in Norway*. The Norwegian Institute of Local History and Local History in Norway, Oslo: NLI. Available at: http://www.lokalhistorie. no/english/local-his.html (accessed: 15.09.2010).

Andersen KV (2006) E-government: Five Key Challenges for Management. *The Electronic Journal of eGovernment* 4(1):1-8.

Andersen KV, Grönlund Å, Moe C E & Sein M K (2005) Introduction to the Special Issue Egovernment in Scandinavia. *Scandinavian Journal of Information Systems* 17(2): 3-10. Available at: https://pdfs.semanticscholar.org/9545/7c9e9f70c7d44dc3582ea0ce96e0df38131f.pdf (accessed: 8.11.2016)

Baker KS & Millerand F (2007) Articulation Work Supporting Infrastructure Design: Coordination, Categorization, and Assessment in Practice. In: *Proceedings of the 40th Hawaii International Conference on System Sciences:* 1-10.

Baker KS, Jackson SJ & Wanetick JR (2005) Strategies supporting heterogeneous data and interdisciplinary collaboration: towards an ocean informatics environment. In: *System Sciences, 2005. HICSS'05. Proceedings of the 38th Annual Hawaii International Conference on*. IEEE: 219b-219b.

Barry A (2001) *Political Machines: Governing a Technological Society.* London: The Athlone Press.

Björgvinsson E (2014) The Making of Cultural Commons: Nasty Old Film Distribution and Funding. In: Ehn P, Nilsson EN & Topgaard R (eds) *Making Futures: Marginal Notes on Innovation, Design, and Democracy*. Cambridge, Massachusetts: MIT Press.

Björgvinsson E, Ehn P & Hillgren PA (2012) Design Things and Design Thinking: Contemporary Participatory Design Challenges. *Design Issues* 28(3): 101-116.

Bossen C & Markussen R (2010) Infrastructuring and Ordering Devices in Health Care: Medication Plans and Practices on a Hospital Ward. *Computer Supported Cooperative Work* 19(6): 615-637.

Botero A, Paterson A, Saad-Sulonen J (eds) (2012) *Towards peer production in public services: Cases from Finland*. Aalto University publication series Crossover 15/2012 Helsinki: Finland.

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

Bowker G, Baker K, Millerand F & Ribes D (2010) Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: Hunsinger J et al. (eds) *International Handbook of Internet Research*. Dordrecht: Springer, 97-117.

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

Bowker G & Star SL (1999) *Sorting Things out Classification and its Consequences.* Cambridge, MA: MIT Press.

Braudel F (1958) La longue duree, Annales E.S.C 4, 725-753

Bruns A (2008) *Blogs Wikipedia, Second Life and Beyond From Production to Produsage*. New York: Peter Lang.

Bødker K, Kensing F & Simonsen J (2004) *Participatory IT Design. Designing for Business and Workplace Realities*. Massachusetts: MIT Press.

COM (2010) 743 Final The European eGovernment Action Plan 2011-2015 Harnessing ICT to promote smart sustainable & innovative Government. Available at: http://eur-lex.europa. eu/LexUriServ/LexUriServdo?uri=COM:2010:0743:FIN:EN:PDF (accessed: 8.11.2016).

COM (2011a) 877 Final Proposal for a Directive of the European Parliament and of the Council Amending Directive 2003/98/Ec on Re-Use of Public Sector Information.

COM (2011b) 882 Final Communication from the Commission to the European Parliament the Council the European Economic and Social Committee and the Committee of the Regions.

Crook E (2007) *Museums and Community Ideals Issues and Challenges*. Milton Park: Routledge.

Crook E (2010) The Politics of Community Heritage: Motivations, Authority and Control. *International Journal of Heritage Studies* 16(1-2): 16-29.

de Jong T& Ferguson-Hessler MGM (1996) Types and Qualities of Knowledge. *Educational Psychologist* 31(2): 105-113.

Directive 2003/98/EC (2003) Directive 2003/98/Ec of the European Parliament and of the Council of November 2003 on Re-Use of Public Sector Information.

Dollar C M (1992) *Archival theory and information technologies: The impact of Information Technologies on Archival Principles and Methods.* University of Macerata, Macerata

Economics and the Common(s) (2013) Economics and the Common(s): From seed form to core paradigm. A report on an international conference on the future of the commons organized by the Heinrich Böll Foundation, Commons Strategies Group, Charles Léopold Mayer Foundation, and Remix the Commons. Available at: https://www.boell.de/sites/default/files/ecc_report_final.pdf (accessed: 8.11.2016).

Edwards P N, Jackson S J, Chalmers MK, Bowker G C, Borgman C L, Ribes D, Burton M & Calvert S (2013) *Knowledge Infrastructures: Intellectual Frameworks and Research Challenges*. Ann Arbor: Deep Blue. Available at: http://hdl handle net/2027 42/97552 (accessed: 10.10.2015).

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

Edwards PN, Jackson S J, Bowker G C, Williams R (2009) Introduction: An Agenda for Infrastructure Studies. Journal of the Association for Information Systems, 10; special issue, 364-374

Ehn P (2008) Participation in Design Things. *Proceedings of Participatory Design Conference 2008,* Indiana Bloomington, USA, September 30-October 4 2008.

Fujimura J (1987) Constructing 'Do-Able' Problems in Cancer Research: Articulating Alignment. *Social Studies of Science* 17: 257-293.

Gauld R Grey A & McComb S (2009) How Responsive Is Egovernment? Evidence from Australia and New Zealand Government. *Information Quarterly* 26(1): 69-74

Grinter R (1996) Supporting Articulation Work Using Software Configuration Managements System. *Computer Supported Collaborative Work: The Journal of Collaborative Computing* 5:447-465.

Greimas AJ & Courtes J (1986) Sémiotique II: Dictionnaire raisonné de la théorie du langage. Hachette, Paris.

Gomart E & Hennion A (1999) A Sociology of Attachment: Music Amateurs, Drug Users. In: Law J & Hassard J (eds) *Actor Network Theory and After.* Oxford: Blackwell, 220-47.

Grönlund Å & Ranerup A (eds) (2001) *Elektronisk Förvaltning, Elektronisk Demokrati. Visioner Verklighet Vidareutveckling.* Lund: Studentlitteratur.

Halonen A (2012) *Being Open About Data Analysis of the UK open data policies and applicability of open data*. The Finnish Institute in London. Available at: http://finnishinstitute.cdn.coucouapp.com/media/W1siZiIsIjIwMTYvMDgvMjUvMDkvMzgvMjEvZGQyOGQ5MTUtY2Q4YS00N2U0LWEyZTAtM2FiOWUxN2M0Njc2L2JlaW5nIG9wZW4gYWJvdXQgZGF0YS5wZGYiXV0/being%20open%20about%20data.pdf?sha=cad600352a6ffb0b (accessed 12.11.2016)

Haraway D (1988) Situated Knowledges: The Science Question in Feminism and the Privilege of Partial Perspective. *Feminist Studies* 14(3): 575-599.

Hennion A (2012) Attachments: A Pragmatist View Of What Holds Us. In: *The First European Pragmatist Conference*, Roma 19-21 September 2012.

Hellberg AS & Hedström K (2014) The Story of the Six Myth of Open Data and Open Government. *Transforming Government: People, Process and Policy* 9(1): 35-51.

Hess C (2008) Mapping the New Commons. In: *Governing Shared Resources: Connecting Local Experience to Global Challenges*. Presented at the 12th Biennial Conference of the International Association for the Study of the Commons. University of Gloucestershire, England. Available at: http://dx.doi.org/10.2139/ssrn.1356835 (accessed: 8.11.2016).

Hine C (2000) *Virtual Ethnography*. London, Thousand Oaks, New Delhi: SAGE.

Hine C (ed) (2005) *Virtual Methods Issues in Social Research on the Internet.* Oxford, New York: BERG.

Hillgren PA, Seravalli A & Emilson A (2011) Prototyping and Infrastructuring in Design for Social Innovation. *CoDesign* 7(3-4): 169-183.

Holdgaard N & Klastrup L (2014) Between Control and Creativity: Challenging Co-creation and Social Media Use in a Museum Context. *Digital Creativity* 25(3): 190-202.

Karasti H (2014) Infrastructuring in Participatory Design. In: *Proceedings of Participatory Design Conference 2014*, Windhoek Namibia 6-10 October 2014.

Karasti H (2001) Increasing sensitivity towards everyday work practice in system design. Dissertation, University of Oulu.

Karasti H (2003) Can film developers be(come) technology developers? Reflections on gendered expertise and participation in system design. In: Mörtberg C, Elovaara P & Lundgren A (eds) *How do we make a difference: information technology, transnational democracy and gender.* Luleå: Luleå University of Technology, 20-49.

Karasti H & Baker K S (2004) Infrastructuring for the Long-Term: Ecological Information Management. In: *Proceedings of HICSS'37*. Hawaii, USA, January 5– 8, 2004, 10.

Karasti H & Syrjänen AL (2004) Artful Infrastructuring in two Cases of Community PD. In: *Proceedings of Participatory Design Conference 2004*, 20-30. Artful Integration: Interweaving Media, Materials and Practices. Toronto, Canada, July 27–31, 2004. ACM Press, 20–30.

Kellogg KC, Orlikowski W J & Yates J (2006) Life in the Trading Zone: Structuring Coordination across Boundaries in Postbureaucratic Organizations. *Organization Science* 17(1): 22-44.

Lathrop D & Ruma L (eds) (2010) *Open Government - Collaboration Transparency and Participation in Practice.* Sebastopol: O´Reilly Media Inc.

Latour B (1999 [2004]) *Politics of Nature: How to Bring the Sciences into Democracy*. Cambridge, MA: Harvard University Press.

Le Dantec C A & DiSalvo C (2013) Infrastructuring and the formation of publics in participatory design. *Social Studies of Science* 43(2): 241-264.

Lessig L (2004) *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York: Penguin Press.

Marres N (2007) The Issues Deserve More Credit: Pragmatist Contributions to the Study of Public Involvement in Controversy. *Social Studies of Science* 37: 759-780

Mörtberg C, Bratteig T, Wagner I, Stuedahl D & Morrison A (2010) Methods that matter in digital design research. In: Wagner, Bratteteig, Stuedahl (eds) *Exploring Digital Design*. Multi-disciplinary design practices. London: Springer verlag, 105-147.

Oomen J & Arroyo L (2011) *Crowdsourcing in the cultural heritage domain: opportunities and challenges*. Paper presented at 5th International Conference on Communities & Technologies in Brisbane, Australia. Available at: http://dl.acm.org/citation.cfm?id=2103373&CFID=864331598&CFTOKEN=15293815 (accessed: 12.11.2016).

Owens T (2013) Digital Cultural Heritage and the Crowd. *Curator: The Museum Journal* 56(1): 121-130.

Parker R & Bradley L (2000) Organizational Culture in the Public Sector: Evidence from Six Organizations. *International Journal of Public Sector Management* 13(2): 125-141.

Perkin C (2010) Beyond the Rhetoric: Negotiating the Politics and Realizing the Potential of Community-Driven Heritage Engagement. *International Journal of Heritage Studies* 16(1): 107-122.

Pipek V & Wulf V (2009) Infrastructuring: Toward an Integrated Perspective on the Design and Use of Information. Technology *Journal of the Association for Information Systems* 10(5): Article 1.

PSI DIRECTIVE 2013/37/EU (2013) DIRECTIVE 2013/37/EU OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information http://ec europa eu/digital-agenda/en/european-legislation-reuse-public-sector-information (accessed 2014-02-17)

Ribes D & Finholt TA (2009) The Long Now of Technology Infrastructure: Articulating Tensions in Development. *Journal of the Association for Information Systems* 10(5): Article 5. Available at: http://aisel.aisnet.org/jais/vol10/iss5/5 (accessed: 8.11.2016).

Runardotter, M (2011) Organizational cooperation for cultural heritage—A viable systems approach. *System Research and Behavioral Science* (28): 77–90. doi:10.1002/sres.1047

Runardotter M, Mirijamdotter A & Mörtberg C (2011) The Changing Nature of Archives – Whose Responsibility? *Electronic Journal on eGovernment* 9(1): 68-78.

Røssaak E (ed) (2011) *The Archive in Motion: New Conceptions of the Archive in Contemporary Thought and New Media Practice*. Oslo: Studies from the National Library of Norway NOVUS, 11-29.

Sefyrin J & Mörtberg C (2009) "We Do Not Talk About This" – Problematical Silences in E-government. *Electronic Journal of eGovernment* 7(3): 259-270.

Schmidt K & Bannon L (1992) Taking CSCW seriously: Supporting Articulation Work. *Computer Supported Cooperative Work* 1(1): 7-40.

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

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

Star S L & Ruhleder K (1994) Steps Towards an Ecology of Infrastructure: Complex Problems in Design and Access for Large-Scale Collaborative Systems. In: *Proceedings of Computer Supported Cooperative Work (CSCW) Conference,* Chapel Hill, North Carolina USA, October 22-26 1994: 253-264.

Star S L & 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.

Stevens M, Flinn A & Shepherd E (2010) New Frameworks for Community Engagement in the Archive Sector: from Handling over to Handing on. *International Journal of Heritage Studies* 16(1): 50-76.

Stoltzfus K (2005) Motivations for Implementing E-Government: An Investigation of the Global Phenomenon. In: *Proceedings of National conference on digital government research,* Atlanta, Georgia: 333-338.

Stuedahl D (2009) Digital Cultural Heritage Engagement - A New Research Field for Ethnology. *Ethnologia Scandinavica* 39: 67-81.

Stuedahl D (2011) Social Media and Community Involvements in Museums. A case study of a local history wiki community. *Nordisk Museologi* 2011(1): 3-14.

Strauss A (1993) *Continual Permutations of Action*. New York: Aldine De Gruyter.

van der Velden M (2013) Decentering Design: Wikipedia and Indigenous Knowledge. *International Journal of Human-Computer Interaction* 29(4): 308-316.

Van Passel E & Rigole J (2014) Fictional institutions and institutional frictions: creative approaches to open GLAMs. *Digital Creativity* 25(3): 203-211.

Vickery G (2011) *Review of Recent Studies on Psi Re-Use and Related Market Developments 2011* Available at: https://www.nsgic.org/public_resources/Vickery.pdf (accessed 31.5.2012)

Waterton C (2010) Experimenting with the Archive: STS-ers as Analysts and Co-constructors of Databases and Other Archival Forms. *Science Technology & Human Values* 35(5): 645-676.

Waterton E (2010) The Advent of New Digital Technologies. *Museum Management and Curatorship* 25(1): 5-11.

Waterton E & Watson S (eds) (2010) *Heritage and community engagement: collaboration or contestation?* Routledge: London.

Notes

1 YEAH Project description Final Feb. 2012
2 Meeting notes 2012-01-18
3 Meeting notes 2012-02-02; Meeting notes 2012-02-20
4 Meeting notes 2012-03-1; Meeting notes 2012-03-26
5 Meeting notes 2012-04-24/25
6 Meeting notes 2012-10-24/25
7 Meeting notes 2012-06-27
8 Meeting notes 2012-09-04
9 Meeting notes 2013-03-25/26
10 The Handbook is available at: http://pure.ltu.se/portal/files/96261604/YEAH_Handbook_ver_1_1_20140506.pdf (accessed: 14.11.2016).

Pictures at an Exhibition – and Beyond. Review of the 'Reset Modernity!' exhibition, 16.04.2016 – 21.08.2016, ZKM, Karlsruhe, Germany.

Endre Dányi

Department of Sociology, Goethe University, Frankfurt am Main, Germany /danyi@em.uni-frankfurt.de

Michaela Spencer

Northern Institute, Charles Darwin University, Darwin, Australia

'Reset Modernity!' is an impressive exhibition curated by Bruno Latour, Martin Guinard-Terrin, Christophe Leclercq, and Donato Ricci at the Zentrum für Kunst und Medientechnologie (ZKM) in Karlsruhe, Germany.[1] It is the third exhibition that Latour has co-curated in this space, (or perhaps being mindful of *Laboratory Life* (Latour & Woolgar 1986), we should say 'this place'). The first two were *Iconoclash* (2002) and *Making Things Public* (2005), both curated with Peter Weibel. This exhibition is also the fourth product of the 'An Inquiry into Modes of Existence' (AIME) project: in 2012 we saw the publication of the AIME book and the website,[2] then in 2013 the staging of a series of workshops, ending with a two-day conference in Paris in 2014. As the fourth instalment, 'Reset Modernity!' exhibits a wide range of images, videos, installations, and texts loosely following the themes of the AIME project, leading the visitor through several truth regimes or modes of existence associated with Western modernity.

Crucial to viewing the exhibition is the 'Field Book', presented as a companion throughout our visit. With text in both German and English, it informs us that the path through the exhibition space is divided into 'six procedures, each allowing for a partial reset [of modernity]'. Attempting to elicit active engagement on the part of the visitor, it notes:

> As the name 'field book' indicates, you are invited to do a bit of research yourself. In each procedure you will find a sort of workplace, called a 'station': this is where you will find more information and where you can discuss the path of the inquiry.

Our review takes at face value the exhibition as an opening onto possibilities for doing further research. At first we had tried to be compliant visitors, following the structure of the exhibition from beginning to end, but it was not long before we found ourselves on a somewhat different trajectory. In the first part of our review we narrate this unplanned journey, which was triggered by our experience of the exhibition. In the second part, we then re-consider the exhibition in light of this journey and where it led us.

Experiencing the Exhibition

What brought us to Karlsruhe was our own collaborative research project titled 'Landscapes of Democracy', which aims to contrast various places and practices associated with democratic politics in Germany (especially Berlin), and Australia (especially Indigenous Australia). 'Knowing landscapes' is one of our central analytic concepts, so while the entry point to the exhibition is very clear, we were immediately drawn to a reprint of a nineteenth century painting of a river enclosure. The item occurs early in the exhibition, as part of 'Procedure A - Relocalizing the Global'. Explaining this, the Field Book notes 'it is important to show the gaps separating the many different instruments and the legions of skilled engineers and scientists. They are those who would need to assemble different viewpoints in order to guide the observer from galaxies to atomic particles' (Latour et al. 2016: 5).

The reproduced painting we are pulled up by is by Caspar David Friedrich, perhaps the most famous German romantic painter, showing a spot by the Elbe near Dresden. Although it was painted in 1831, the scene reminds us of a digital photograph. Not only the colours, which capture an exceptionally calm, beautiful moment right after sunset, but also the way the perspective curves, creating a fisheye lens effect. The Field Book text tells us that the curving perspective makes this painting quite special, as it generates an impossible vantage point: we are, it seems, at once in, on, and above a particular 'spot', the river itself constituting the background. We are almost floating in the warm air; we can almost smell the sour mud; we can almost hear the birds and the bugs, circling excitedly above the water before the dark settles.

Then, on the same wall, there is an almost identical copy of Friedrich's painting. This print is black and white. In this contrast alone already some of the beauty is gone. But, as the Field Book points out, there is more to be noticed. The main difference lies in the straightened or flattened perspective. The horizon is straight, the proportions are 'correct', and we are no longer floating between heaven and earth. Rather, we are forced to stand on the ground with both of our feet, and observe the place that once had such a strong effect on Friedrich. What could be the purpose of this imperfect copy? It is as if the two images were to illustrate Lorraine Daston and Peter Galison's (1992) argument about different understandings of objectivity: the first tries to stay faithful to the skilled and gifted observer's view, while the second tries to stay faithful to the place itself. In this second objectivity the river, the trees, the clouds, the mud flats—they all become parts of an invisible inventory amenable to management by science.

We spend an unusually long time discussing these two images. Pulled so deeply into the contrast of this exhibit, it is difficult to do justice to the rest. Nevertheless, we do make an effort to return to the start and are happy to discover subtle references to early STS works along our way through subsequent procedures. We are shocked by five identical images of a quarry, the making of which required platinum extracted from one ton of ore. We are amused by a robotic arm that draws us an up-to-date map of the glacier that constitutes the natural border between Austria and Italy. And we are impressed by an installation that shows a speech by U.S. president Barack Obama at a Methodist church meeting, along with the audience's reactions and a conversation analytical transcript. Our passage though the exhibition however does not comply with the elaborate instructions given in the Field Book; we fail to fully connect the partial resets of each of Procedures A-F. Our discussions keep coming back to the two depictions of the river Elbe.

Leaving the exhibition we go to the ZKM bookshop. We notice there is a thick catalogue associated with the exhibition, edited by Bruno Latour and Christophe Leclercq (2016). One of the chapters is dedicated to Friedrich and his 'Large enclosure'. The author, art historian Joseph Leo Koerner (2016), explains that the spot in Friedrich's painting is called Ostragehege, located not too far from Dresden city centre. Friedrich used to live nearby and knew the river well. His painting is clearly a testimony of his love for the place. The other painting, we learn, is a catalogue entry prepared by Johann Philipp Veith for the Art Association of Saxony. It is not an artwork *per se*, but a record that documents the Association's acquisition of Friedrich's original in 1832. To our great surprise, in Koerner's chapter we discover a third version of the painting by South African artist William Kentridge. This version, made in 2014, is a charcoal drawing based on the 'Large enclosure': the colours are still missing but this time the curves are back. The paper Kentridge uses for this drawing apparently comes from the ledger of the 1906 cash book of Johannesburg's Central Administration Mine. This unusual medium suggests to us that the river and its curves are not the whole picture, since under the surface there are even larger forces at play – the forces of capitalist development.

As we leave the bookshop, we wonder about the three images and how they relate. To use the language of the AIME project, they might be recognised as emblems of three modes of existence: art [FIC], science [REF] and politics [POL]. Multiple truth claims about a singular place: is this the inquiry we need to pursue to reset modernity? To recognise multiple, distinct and inter-relatable ways of making truth claims? Somehow we feel something is still missing. What is it? Experience of the embodied here and now? Taking the Field Book's encouragement to do a bit of research ourselves, we decide to travel to Dresden to find the large enclosure. Our idea is not to overwrite the 'Reset Modernity!' exhibition, but to playfully extend it beyond the walls of the ZKM.

* * *

Getting to Dresden is not difficult (even though one of us had to travel half the globe in order to even be in the vicinity), but how do we find the place depicted by the three images of Friedrich, Veith, and Kentridge? We know the area is called 'Ostragehege': it lies about 2 kilometres out of the city centre, on the south bank of the Elbe, west of a bridge called Marienbrücke. According to the German version of the relevant Wikipedia entry,[3] in Friedrich's time it was part of a larger floodplain. In the late nineteenth and early twentieth centuries, the flow of the river was regulated, so that the area could be further developed. Some of the subsequent development projects were the establishment of a massive slaughterhouse, a harbour, and several sports clubs. These days, the Ostragehege is better known as a protected natural habitat of about 35 species of fish – at least this is what our quick online research tells us.

We arrive at Dresden around noon. From the main station we go straight to the Marienbrücke and continue our extended exhibition tour westward along the river. We walk for several hours in the summer heat, stopping from time to time at small inlets and beaches to skip stones, watch the fish, smell the mud, and listen to the birds and the bugs. We also take some time to sit down, chat, and make a few notes and drawings of our own. However, as the evening comes closer, there is a lingering sense that while we have made the effort to travel here, we have not succeeded in finding what we were looking for: the spot that Friedrich loved – or so the story goes.

If we wanted to experience Friedrich's sublime nature, perhaps it would have been better to go to the Neue Galerie in the city centre, where the original of the 'Large enclosure' is on display. We suspect art museums might be more appropriate sites to engage with nineteenth century landscape paintings than the places they depict. What then about the corrected reality of Veith's almost identical copy? The rationality of the straightened lines and perspectives that was once associated with a landscape well known to science, is also difficult to celebrate in the Ostragehege. As it turns out, the area was a primary target during the bombing of Dresden in 1945; the slaughterhouse constructed in the early twentieth century was the site where Kurt Vonnegut was imprisoned in the last days of the Second World War. These days, there are hardly any visible traces left of the devastation described in *Slaughterhouse-Five* (1969), but it is also impossible to un-see them as we walk past the remnants of the huge industrial complex. It is tempting to claim we can sense the larger forces of capitalist development Kentridge captures so well in his drawings (see a recent exhibition at the Martin Gropius Bau in Berlin),[4] but we feel ill-equipped to access them. Our time is running out, as we need to catch the last train to Berlin. We are tired and confused. We somehow feel we had a better understanding of the place *before* we travelled here.

Reviewing the Journey

What might we make of this journey? The promise of Friedrich's, Veith's, and Kentridge's images as they were displayed in Latour et al.'s exhibition and catalogue is that they sensitise us to multiple truth regimes – the felicities of divergent modes of existence associated with art, science, and politics, among others – held together by the work of a curatorial team in a singular place. Staying true to these multiple truth regimes, so to speak, and finding better ways of holding them together, might help us reorient ourselves in a world where our modern devices and institutions seem to be less and less adequate. However, when visiting the Ostragehege, the clashing and melding of multiple truth regimes was not what we found. Rather, it was the inchoate happening of a place, the experience of a place, which was radically unknown to us. It pushed back and exposed the naivety of our initial assumptions.

We felt we needed to leave the exhibition, and travel beyond Karlsruhe, if our review of the exhibition was to take seriously the moves it had proposed. Our journey, in turn, showed that engaging with this place as a 'knowing landscape' would require far more than what might be called 'epistemic tourism': a quick trip to Dresden and an afternoon's stroll along the Elbe. It also revealed that collectively resetting modernity would require more than six sequential procedures that identify, and therefore permit partial shifts, in defined modes of modern epistemic practices. Along our journey, we failed to reset modernity because we were unprepared for the pushing back of the place that necessitates us to also be knowingly involved in at least some of the stories that might be told about it.

Does our failure shed bad light over 'Reset Modernity!'? It depends on how we understand the purpose of the exhibition. If it is to offer visitors a guide that effortlessly extends space beyond the walls of the ZKM, rendering every place potentially as reset-able as a museum or a laboratory 'no-place', then we suspect our confusion as trained STS scholars is collective. If, however, the purpose is to encourage visitors to take the insights of forty years of STS to new places, to send them on a walk and make them sweat, then our unplanned journey shows that the exhibition works. We are still discussing, and we are grateful for the experience.

References

Daston L & Galison P (1992) The image of objectivity, *Representations*, 40: 81-128.

Koerner J L (2016) Caspar David Friedrich: Earth Life Art, in: Latour B & Leclercq C (eds.) *Reset Modernity!* Cambridge: The MIT Press, 68-73.

Latour B & Leclercq C (2016) *Reset modernity!* Cambridge: The MIT Press.

Latour B, Guinard-Terrin M, Leclercq C, Jansky C & Havemann U (2016) *Reset Modernity! – The Field Book*. Karlsruhe: ZKM.

Latour B (2013) *An Inquiry into Modes of Existence*. Cambridge: Harvard University Press.

Latour B & Weibel P (2005) *Making Things Public: Atmospheres of Democracy*. Cambridge: The MIT Press.

Latour B & Weibel P (2002) *Iconoclash*. Cambridge: The MIT Press.

Latour B & Woolgar S (1986) *Laboratory Life: The Construction of Scientific Facts*. Princeton: Princeton University Press.

Vonnegut K (1969) *Slaughterhouse-five; or, The children's crusade, a duty-dance with death.* New York: Delacorte Press.

Notes

1 http://zkm.de/en/event/2016/04/globale-reset-modernity (accessed:22.8.2016)
2 http://modesofexistence.org (accessed: 22.8.2016)
3 https://de.wikipedia.org/wiki/Ostragehege (accessed: 22.8. 2016)
4 http://www.berlinerfestspiele.de/en/aktuell/festivals/foreign_affairs/fa16_kentridge/fa16_kentridge_1.php (accessed: 22.8. 2016)

Geoffrey C. Bowker, Stefan Timmermans, Adele E. Clarke, & Ellen Balka (eds) (2015) Boundary Objects and Beyond. Working with Leigh Star. Cambridge MA, London: MIT Press. 548 pages. ISBN: 978-0-262-02974-2

Attila Bruni
attila.bruni@unitn.it

I knew when I agreed to review this text that it would be a challenge. Leigh Star (although for some she was Susan, as recalled by John Leslie King - chapter 17) was one of the most influential scholars of science and technology studies of the past thirty years, and her contribution has marked (and still inspires) an entire field of research. Moreover, she was renowned for the calmness and serenity that she conveyed, the poems that she inserted in her writings and speeches, and the almost whispered – but incisive – tone of voice with which she used to speak. That is to say, it is almost impossible to speak of Leigh Star's work without speaking of Leigh herself. Hence reviewing the book that originates from the conference held in her honour at the University of California, San Francisco, in September 2011 carries the serious risk of producing nothing more than an apologia for her work and her person. To avoid this risk, therefore, I promise myself and the reader not to address Leigh Star's research in this review, even less to describe her personality, but instead to concentrate entirely on the book.

From a purely material point of view, the book is big, heavy and it has a somehow unusual page format which makes it larger than the 'normal' books you have in your library. The text (articulated in four sections: Ecologies of Knowledge; Boundary Objects; Marginalities and Suffering; Infrastructure) consists of collection of writings by Star accompanied by contributions from coauthors/colleagues/friends (the three categories often overlap). The contributions are often outright research papers which reflect on Star's concepts and approach, or which show the efficacy and practical relevance of her thought in light of various research examples. Others are instead more narrative in form, as in the case of the 'epistolary' text in which Brian Cantwell Smith (chapter 9) draws on the concept of boundary object to extend the discussion to what counts as an object, and what is implied ontologically by looking at the world in terms of objects. Or in the case of Eevi Beck, who presents a patchwork of different texts and rhetoric styles in order to thematize "vague areas of science; where the solidity of theories seems to peter out; where scientists emerge as eminently human in our splendor and our fallibility" (chapter 22: 436). Or in the case of Nina Wakeford (chapter 3), who discusses Leigh's queerness by considering a working paper written by Star for a feminist conference at Aarhus University in 1994. Wakeford then illustrates the making of a four-minute digital video (for the conference in Leigh's honor) and reflects on how to challenge and innovate the methodological repertoire of STS.

The complete version of Star's paper appears some chapters later (chapter 6). As I read it (I had never heard mention of it before), I was reminded of my reactions on first reading Star on the occasion of my degree thesis, when I stumbled upon the famous 'onion paper'. Magnificent, I thought. Incomparable, splendid, as well as other adjectives indicating beauty or exceeding all expectations. In this paper Leigh transports the reader to the San Francisco of the 1970s and the do-it-yourself technologies and practices for artificial insemination in the lesbian community. She then addresses the question of feminism as a method, and of methodology as an experiential (and political) form of ordering practice. I shall not go into details on the chapter so as not to deprive readers of the pleasure and surprise that reading it evokes, but I confess that I bit my lip on reading that "the trajectory of learning becomes a series of encounters with the objects of practice in the community" (chapter 6: 154); a reflection which I thought constituted the original feature of a paper of mine published in 2001 (sic!).

In fact, on reading the entire book one gains the distinct impression that Leigh Star anticipated the main themes that animate the current debate not only in STS, but more generally in the social sciences. I think three examples are sufficient.

The first is the attention paid to the role played by classifications and infrastructures in the contemporary world, and to the idea itself of infrastructure as a relational concept. For example, if we consider the programme of the last 4S/EASST conference, we find that the word 'infrastructure' is one of the most recurrent (to be exact, it appears in the titles of 61 papers, and in the names of 6 different tracks); this year 4S has established an 'STS Infrastructure Award'; and if we review the last five years of the main STS journals (such as *Science, Technology and Human Values*, *Science and Technology Studies*, and *Science, Technology and Society*), the theme of infrastructure is ubiquitous. If we then move from STS journals to more generalist ones (such as *Theory, Culture and Society*, or *Sociology*), the result does not change: indeed, the numbers increase. Although these are mere numerical indicators (and Star's research demonstrates how many different forms of reductionism and standardization lie concealed behind numbers), I believe that it can be argued that the theme of infrastructure is today of key importance for those who deal with STS and/or try to read the 'social' through an STS lens (not by chance, this same book is part of the MIT 'Infrastructures Series'). Of course, this also applies to the attention devoted to infrastructure by diverse STS scholars (first of all, Geoffrey Bowker); but also in this case I feel sufficiently confident in stating that "Steps toward an ecology of infrastructure: design and access for large information spaces" and "The ethnography of infrastructure" (both in the text) are seminal articles for anyone dealing with this issue.

The second theme is instead linked to the relationship between the visible and invisible, and in particular, to the articulation work that makes it possible to hold together the two poles of what (like so many other oppositions and clashes) should be conceived as a continuum rather than a dichotomy. This is the issue that more than any other ties Star's approach to symbolic interactionism and makes social practice the prime unit of analysis with which to determine what people actually do when they work; as well as how much and what type of work people must perform 'covertly' before they can devote themselves to 'real-and-proper' work. In other words, what activities count as 'work' and what activities (or actors) are instead removed.

In/visible work has been a core theme for diverse bodies of inquiry (CSCW, HCI, organization studies, among others), becoming a concept almost taken for granted in contemporary STS. This has led to a paradoxical consequence and to a further variant of invisible work, which Star and Strauss mention only in passing, but which Kjeld Schmidt well evidences in his contribution (chapter 18: 346): "the elimination of work from the agenda of respectable intellectual interests". Again, I think it suffices to look at the programme of the last 4S/EASST conference to grasp the importance assumed by the theme of work in contemporary STS: of the 185 tracks on the conference programme, not even one centres specifically on the work/organization dynamics that characterize technoscientific contexts. And also when considering the titles of the individual papers, the word 'work' appears much more often as an item comprised in other words (primarily 'network' and 'framework'), not as a word in its own right. As a researcher interested in action of technologies in organizational contexts, I cannot but endorse the exhortation with which Schmidt closes his chapter (18: 349): "a fine way to honor Leigh's work would be to strive to make work –

the skilful work practices and the problems and concerns of ordinary workers – visible again: make it a legitimate and respectable topic of study". This is also because, one might add, just as organization studies have found in STS a new conceptual apparatus with which to read phenomena related to work and organizational processes, so STS may find in contemporary studies on work and organizational practices stimuli and suggestions with which to enrich and renew their interpretative instruments.

It is well known, however (and emphasised on several occasions in the book) that Star regarded the relationship between the visible and invisible as only a starting point from which to address broader issues and dynamics involved in the construction of borders, margins, differences and inequalities. The definition of categories is a political issue, especially in a world where knowledge is increasingly shared among humans, machines and classification standards. As well as in Star's writings, clear evidence of this process is provided in the book's chapters by Gail Hornstein (chapter 14) and Jutta Weber and Cheris Kramarae (chapter 15). Although they deal with very different research objects (respectively, the construction and role of categories in psychiatry, and the categorization of 'civilians' and 'victims' in the war currently being waged in Pakistan by the United States), both contributions testify to the inspirational impact of Star's ideas. They do so in line with the interactionist and 'grounded' approach to the development of sensitizing concepts, able to stimulate the imagination and creativity of the researchers who want to engage with them.

As transpires from several of Star's writings (my favourite has always been, and remains so upon re-reading it, "Living Grounded Theory: cognitive and emotional forms of pragmatism" - chapter 5), her relationship with concepts and ideas was physical, corporeal, and embodied. This brings us to the third great current theme anticipated by Star: affect.

As pointed out by Adele Clarke on discussing different forms of anticipation work (chapter 4: 105), perhaps it was no accident that, as president of 4S, Leigh chose the theme of "Silence, Suffering and Survival" for the annual meeting of the Society. We must therefore bear in mind that Star's vocabulary consisted not only of concepts such as boundary object, infrastructure, and in/visible work, but also of truth, spirituality and hope. These concepts are not to be interpreted in 'modern' terms but rather in their post-structuralist sense as spaces-between, occasions that prevent *a priori* definition of being and the development of events, and interactions. These are concepts which, as noted by Maria Puig de la Bellacasa (in a chapter that I found so beautiful and inspiring that I wish I had written it myself), perfectly matches the ecological formulation of Star's thought. "An ecology (...) evokes a site of intensities, synergies, and symbiotic processes within relational compounds. Ecological circulation functions in cyclic interdependent ways rather than by extension. (...) Ecological thinking is attentive to the capacity of relation-creation, to how different beings affect each other, to what they do to each other, the internal 'poiesis' of a particular configuration." (chapter 2: 53).

Also the famous *cui bono?* underpinning Star's work since publication of *Ecologies of Knowledge* (here rightly re-proposed as the book's opening chapter) is a question that does not have the ambition to balance the benefits that some can draw from situations at the expense of others, but to make visible what is hidden, repressed or silenced. Reality is not a zero-sum game. Hence, instead of trying to establish equivalences, it may be more sensible (especially from a pragmatic point of view) to state differences and incommensurabilities.

In conclusion, I believe the book succeeds in the difficult task of paying homage to Star's thought and person, outlining various lenses through which to read her work. For my part, I hope that I have kept the promise made at the beginning of this review and been able to arouse the reader's curiosity in this book, which I believe will give much pleasure to both those who already know Leigh Star and those who have not had the fortune of meeting her work. And if I have dwelt too long on Leigh Star and the cultural (even before theoretical or empirical) contribution made by her work, I apologize, but it is due to a question of affect.

Science & Technology Studies

Volume 29, Issue 4, 2016

Guest editorial

Articles

Reviews