

Data Platforms and Cities

Anders Blok

*SODAS University
of Copenhagen (DK)*

**Antoine
Courmont**

Sciences PO, Paris (FR)

Rolien Hoyng

*The Chinese University of
Hong Kong (CN)*

Clément Marquet

Télécom Paris Tech (FR)

Kelton Minor

*SODAS University
of Copenhagen (DK)*

Christian Nold

*University College
London (UK)*

Meg Young

*University of
Washington (US)*

Abstract: This section offers a series of joint reflections on (open) data platform from a variety of cases, from cycling, traffic and mapping to activism, environment and data brokering. Data platforms play a key role in contemporary urban governance. Linked to open data initiatives, such platforms are often proposed as both mechanisms for enhancing the accountability of administrations and performing as sites for 'bottom-up' digital invention. Such promises of smooth flows of data, however, rarely materialise unproblematically. The development of data platforms is always situated in legal and administrative cultures, databases are often built according to the standards of existing digital ecologies, access always involves processes of social negotiation, and interfaces (such as sensors) may become objects of public contestation. The following contributions explore the contested and mutable character of open data platforms as part of heterogeneous publics and trace the pathways of data through different knowledge, skills, public and private configurations. They also reflect on the value of STS approaches to highlight issues and tensions as well as to shape design and governance.

Keywords: data platforms; data labour and reuse; environmental sensors; urban governance; transparency.

Submitted: September 20, 2017 – **Accepted:** April 30, 2017

Corresponding author: Anders Blok, Department of Sociology, University of Copenhagen, Øster Farimagsgade 5, Postboks 2099, 1353 København K, CSS - Building: 16.1.58. Email: abl@soc.ku.dk

Towards Data-driven Urban STS?

Anders Blok and Kelton Minor

Whenever new and power-laden techno-political projects emerge, science and technology studies (STS) should consider itself invited to the fray. In recent years, visions and materialities of data-driven (so-called) ‘smart urbanism’ have come to constitute such a project, with consequences petit and profound for urban governance around the world. Against a backdrop of increased datafication, whereby daily personal routines in the city now generate digital traces in granular detail, public and private urban actors seek to harvest, process, refine and serve up civic data to power their decision-making processes. As STS researchers, it is tempting (to say the least!) in this situation to dust off still-valid critiques of technological determinism and to re-deploy long-standing analytical commitments to socio-technical contingency in city-making, whether of the urban technological frames, politics of urban design or heterogeneous urban assemblage variety (see Farias and Blok 2017). However, what does one do upon realizing that socio-technical contingency is just what urban authorities are looking to *harness* and *extend* through the deployment of new data formats, techniques and infrastructures embedded into the urban fabric?

In this short reflection piece, we want to deploy a local and perhaps parochial example, drawn from our own recent research experience, in order to raise some more general questions about the stakes of STS in data-driven urban governance. The example in focus pertains to a set of urban ambitions, coordinated in and around the municipality of Copenhagen, Denmark, to extend its commitment to cycling as a climate-friendly mode of transportation via the ‘smarter’ planning capacities conveyed by ‘bigger and better’ data on bicycle mobilities (see Boellstorff and Maurer 2015). More precisely, we want to dwell on a specific participatory event, in which planners, consultants, businesses and researchers from several European cities – ourselves included¹ – were invited to assist the Copenhagen municipality in thinking through its data-infrastructure options and imagining, purportedly, innovative solutions.

Two conditioning parameters of this event immediately stand out. First, the way data and its promises are harnessed and channeled in this setting must be seen as responding to a quite *specific* situation of urban governance in Copenhagen, a city keen to extend its transnational ‘front-runner’ position in domains of bicycle infrastructures, in particular, and urban greening and low-carbon transition more generally (see Blok 2012). In this

¹ In actual fact, only one of this text’s authors (Kelton) participated in the event. While the intricacies of our own research trajectory matter to our story, and will be briefly recounted later in the text, for the sake of convenience – and to convey a point about positionality – we write here mostly in the homogenized voice of a ‘we’.

sense, the event exemplifies how capacities of data to exert effects within urban governance are likely to be strategically co-shaped by a whole range of situated urban realities, interests and trajectories, which come to be latched onto and nested within each other. Indeed, a long history of STS reflection on infrastructures (e.g. Star 1999) would lead one to expect as much: new data infrastructures are perhaps less what ‘drives’ urban governance as what may come to exert effects in wider, more distributed and more layered assemblages of urban techno-politics. In a city like Copenhagen, such data assemblages have latched onto bicycle, low-carbon and other existing urban infrastructural projects.

Second, as should be clear, our own situated format of engagement with this setting, as researchers working on data and urban-related issues at the University of Copenhagen, is one of interiority and participation rather than external observation. To start with, this is not so much a matter of us, as ‘proto-STs’ urban researchers, striving to make a reflexive point.² Rather, it is more about how our own everyday research trajectories have constituted us, in the midst of doing other things and being ‘otherwise engaged’ (Harvey and Knox 2008), as now belonging to the diverse field of ‘knowledge-based stakeholders’ with something to contribute to the ‘technical’ side of the Copenhagen event. More generally, we might say, it has to do with how specific forms of interdisciplinary research, in one capacity or another, are *already* integral to the ongoing ‘infrastructuring’ (Dantec and DiSalvo 2013) of such data relations in the service of local urban change and emergent publics.

The point we wish to make on this basis is less one of action-oriented STS being involved in self-conscious intervention in this field of urban practice, nor one of the performativity of STS across domains of techno-scientific politics in general (see e.g. Zuiderent-Jerak and Jensen 2007). It is a much more modest and situated point, related to a reflection on how it is that our own collective research trajectory has come to be relevant to other actors, with other agendas and concerns, in this particular urban milieu (as Isabelle Stengers would term it) – and conversely, how this milieu has come to be relevant to our research (see Savransky 2016). As we will suggest, this event of relevance seems to us to pertain to the *specific* ways in which technical, political and ethical aspects co-implicate each other in this milieu of data-driven urban governance, with implications also, we argue, for how STS might envisage its own stakes in it.

To briefly set the stage: the participatory event in question, called the “Big Data for Cyclists Workshop”, took place in the House of Innovation in Copenhagen on February 15 2017, under the joint auspices of the municipality’s Technical and Environmental Administration and the so-called Climate-KIC Nordic network, a public-private innovation partnership

² We return towards the end to the notion of ‘proto-STs’. Suffice to say that it is meant here to hint at the ambiguities of knowing, on our part, when ‘STs’ starts and stops in an interdisciplinary research setting such as our own.

sponsored partly by the European Commission. Evidently, in a context of mushrooming ‘living laboratories’ for urban sustainability transition (Evans and Karvonen 2011), and the way these intersect with location-sensitive data technologies for the production of ‘speculative’ urban futures (Leszczynski 2016) and the participatory design of new public ‘things’ (Dantec and DiSalvo 2013), this workshop is easily recognizable as one version of a more general pattern. What interests us here are some details of the staging of data promises during the workshop.

From the outset, municipal organizers were adamant to frame the workshop around an unfortunate asymmetry of traffic planning: whereas Copenhagen planners are in possession of fairly detailed locational data on car- and bus-based mobilities in the city, no comparable data exists for those roughly 41% of trips undertaken by bicycle. Stated otherwise in the city’s supporting information brief, while sensors around the city count aggregate bicycle numbers, this data allows for no extrapolation on routes and trip times needed by the city to respond to “congestion, accidents, stalled cars/lorries parked on bike lanes, road works exceeding their permits, debris, and unplanned events” that can inconvenience “tens of thousands of cyclists” that use the city’s central corridors each day. In the face of strong political commitments to increase the share of bicycle-based traffic to 50% by 2025, planners thus face an obvious challenge: how to optimize interventions aimed to improve bicycle infrastructures – and, correlatively, the attractiveness of choosing bikes over cars – with little solid information on current bicycle practices.

In short, the promises of data staged in the workshop were politically infused from the start, framed within a specific narrative of low-carbon transition in Copenhagen traffic planning. Notably, however, the potential perils of privacy violations were absent from the organizers’ stated list of criteria by which ideas solicited from external stakeholders would be judged. Indeed, to help solve this planning-based challenge through ‘bold’ and ‘radically innovative’ uses of urban data was the very mandate through which the organizers elicited data experts (ourselves included) for participation in the workshop.

Concretely, prior to the workshop, each participant had to submit a brief pitch of her or his prototype ‘big data for cyclists’ idea. At the beginning of the workshop, these pitches were then enacted by their creator(s) for an audience of both in-person attendees along with remote viewers watching via video links from a small constellation of universities and workplaces scattered across Europe. Towards the end of the day, a judging panel – consisting of, among others, a former Danish minister of traffic and the Dutch ambassador to Denmark (!) – selected three prototype ideas for further concept development.³

³ We mention this to signal how the workshop was also trans-local in ways reflective of specific and competitive urban geographies of ‘advanced’ bicycle infrastructures, such as those found in major Dutch cities.

More of a platform for competition than collaboration, the workshop floor was thus a bazaar of self-contained data-promises rather than a space for creative cross-fertilization and experimentation (contrast Perng, Kitchin and Evans 2016). This tension between modes of engagement was further enacted when the event organizers instructed participants (ourselves included) to physically root themselves in different parts of the room to develop their ideas independently. In our experience, this physical displacement and separation formed individual islands of interest visible on the floor and constrained the melding of isolated concepts – an observation also duly noted by one of the judges at the event’s conclusion. Hence, while participatory in name, the workshop enacted its own specific modes of (non-)cooperation.

The fact that we found ourselves in this workshop situation speaks to a certain intersection with our own collective research trajectory – albeit, like much else in this rather heterogeneous setting, that the connection is orthogonal and indirect, rather than fully cooperative (so to speak). For a number of years, we have both partaken in a large-scale interdisciplinary research project known as the Copenhagen Social Networks Study, involving anthropologists, economists, philosophers, physicists, psychologists, sociologists and others.⁴ Set up via a self-built data infrastructure, the project deploys mobile phones as devices for studying social networks – via call and SMS logs, Bluetooth and Wi-Fi records, GPS coordinates and so on – among a freshman class of approximately 800 engineering students at the Danish Technical University (DTU), located north of Copenhagen. As such, it joins the emerging frontier of digital trace-based computational social science, while at the same time experimenting ethnographically and otherwise with the many ethical and political questions thereby opened up (see Blok and Pedersen 2014).

Accordingly, while the Study branches off in many different substantive directions, one recurrent theme reverberating through our interdisciplinary dialogues has taken the shape of a participatory, ‘proto-STs’ interest in the infrastructural technicalities and the political ramifications of data ethics. From the very construction of privacy-protecting databases and data practices at DTU, data ethics has increasingly and recursively become a research topic in its own right for members of our group, as we have sought also to build relations beyond our academic platform. More than this, an ability to speak credibly on data ethics issues has emerged as something of a *resource* for us, as we have realized the sheer salience of such issues amongst everyone from architects to municipal planners. In short, as a team, we have gradually come to inquire into what an ethical data infrastructure might be and what it may become.

⁴ Over the years, the project has carried several names, including SensibleDTU and Social Fabric. Most recently, it forms the backbone of the Copenhagen Center for Social Data Science (SODAS), where this text hails from.

Set within such a generative trajectory, the big data idea submitted for the bicycle workshop by Kelton Minor (co-author of this text), on behalf of our research team, might be said to reflect a collective realization, very much along STS lines, as to the inseparability of technical, ethical and political dimensions when dealing with any digital data infrastructure. The idea, in essence, would be to build on lessons learned during the DTU study on how to use mobile phone-based GPS and Wi-Fi traces to trace bodily mobility routes in sufficiently granular ways to also infer the means of transportation (i.e. bicycles). Such database building would be enabled, in turn, via a two-tiered infrastructure, one based on ‘passively’ recording anonymized, randomized and encrypted mobile device IDs and the other based on ‘actively’ eliciting civic data donations from citizens via an opt-in mobile app as an emergent form of urban volunteerism. Together, the two data-eliciting techniques would provide a proprioceptive picture of aggregated bicycle activity patterns in order to assist the city to sense how citizens are using its extensions (and the hurdles they encounter in doing so).

Through the app-based data donations, this approach activates an element of citizen science, thus working towards engaging urban data citizens and publics into the forging of the algorithmic bicycle city (Paulos, Honicky and Hooker 2009). Conversely, as is immediately obvious, the key to the passive recording of mobile device data and its further processing for planning purposes is how this infrastructure would deal with privacy issues and concerns given the granularity of the data. Here, automated processes of de-identification and randomization ensures that a high level of privacy protection is, so to speak, *built into* the design of the data infrastructure from the outset. Figuratively speaking, the infrastructure works as a community garden-like data commons: while data remains only personally accessible for each individual owner’s ‘plot’ (via app-based data feedbacks), the municipality attains an in-principle de-identified, aggregate overview of the entire ‘garden’ (the city bicycle infrastructure).

From participating in the workshop, it became clear that such a striving for ethically sensitive data infrastructures is by no means a foregone conclusion: other proposals, coming from private data consultancies, would for instance build on face recognition and re-identification techniques from a network of local cameras at traffic-intersections, with little attention to issues of data storage and potential misuse of the powerful responsibilities associated with the capacity to re-identify individuals. Indeed, as noted, privacy protection was initially *not* featured among the criteria of judgment in the workshop competition – something we called attention to during the workshop, as the organizers presented the criteria. On the other hand, when alerted to the latent issues, planners, judges and others proved susceptible to their importance, to the point of this becoming a stated reason as to why our idea was selected for further concept development. In this sense, the workshop itself emerged as a kind of ‘proto-STs’ event, in which the co-shaping of data techniques with ethical concerns came to be partially recognized and embedded into the city’s ongoing proposal formation.

The fate of our prototype big data-for-cyclists idea remains yet to be decided within the involved public-private settings of Copenhagen's governmental entities. As such, it is once again in the hands of more powerful others, pursuing mixed agendas and oriented to additional rationales beyond just our collective research intentions and values enacted in the Copenhagen Social Networks Study. Likewise, it remains to be seen what kind of relevance this foray into the milieus of data-driven urban governance will attain in our future research endeavors (this short text constituting of course only a preliminary start to be sure). So far, we count it as confirmation on our part that a certain sensibility to the contingent co-shaping of the material and the social, the technical and the ethical, within data infrastructures is fast becoming not only the *topic* of urban STS, but also – and this is the point we have wanted to make – a potential *resource* for its further development.

As such, while acknowledging the risk of parochialism and self-indulgence, we have attempted in this short piece to deploy our own situated experiences in the service of perhaps eliciting something critically general: what might a data-driven urban STS come to look like? In this sense, we have attempted to use the Copenhagen workshop also metonymically, as a placeholder for all those relatively underdetermined, inquiry-conducive and awkwardly engaged encounters and distributed spaces that are *also* part of the power-laden and otherwise over-coded landscape of smart urbanism and data governance. We further suggest that, far from being marginal to such spaces and encounters, STS insights into socio-technical contingency might be seen as entirely integral to them – provided, that is, that we as STS researchers are willing to have our 'proto-STS' insights shared across more heterogeneous assemblages of interdisciplinary relations and to recognize their localized embedding in specific urban contexts.⁵

If data is always a contingent socio-technical relation, then a data-driven urban STS might be that endeavor which takes such relations and their urban infiltrations as its own starting and ending point for research, experimentation and critique. Like other forms of digitally informed research, it would constitute an 'interface method' (Marres 2017), shaping up in an as-yet indistinct space of interdisciplinary and extra-academic engagements in the city. Along the way, it will have to come to terms with novel entities and relations, including those emerging urban data publics that remain for us so far only on paper. Plenty of scope persists, then, for improving on this first beta version of data-driven urban STS. We hope other passengers, and drivers, will want to join in the fray to explore the epistemic, technical, ethical and political ramifications that encompass this incipient crossroads.

⁵ As should have become clear, this text is itself the product of such an interdisciplinary writing encounter, as it reflects the joint research commitments of two authors who are *also* otherwise engaged with data and the city.

References

- Boellstorff, T. and Maurer, B. (eds.) (2015) *Data, Now Bigger and Better!*, Chicago, Prickly Paradigm Press.
- Blok, A. (2012) *Greening cosmopolitan urbanism? On the transnational mobility of low-carbon formats in Northern European and East Asian cities*, in “Environment and Planning A”, 44 (10), pp. 2327-2343.
- Blok, A. and Pedersen, M.A. (2014) *Complementary social science? Quali-quantitative experiments in a big data world*, in “Big Data and Society”, 1(2), pp 1-6.
- Dantec, C.A.L. and DiSalvo, C. (2013) *Infrastructuring and the formation of publics in participatory design*, in “Social Studies of Science”, 43(2), pp. 241-264.
- Evans, J. and Karvonen, A. (2011) *Living laboratories for sustainability: exploring the politics and epistemology of urban transition*, in H. Bulkeley et al. (eds.) *Cities and Low-Carbon Transition*, London, Routledge, pp. 126-141.
- Farias, I. and Blok, A. (2017) *STS in the city*, in U. Felt et al. (eds.) *The Handbook of Science and Technology Studies* (Fourth Edition), Cambridge, MA, The MIT Press, pp. 555-581.
- Harvey, P. and Knox, H. (2008) “*Otherwise engaged*”: *Culture, deviance and the quest for connectivity through road construction*, in “Journal of Cultural Economy”, 1(1), pp. 79-92.
- Leszczynski, A. (2016) *Speculative futures: Cities, data, and governance beyond smart urbanism*, in “Environment and Planning A”, 48(9), pp. 1691-1708.
- Marres, N. (2017) *Digital Sociology: The Reinvention of Social Research*, London, Wiley.
- Paulos, E., Honicky, R. and Hooker, B. (2009) *Citizen science: Enabling participatory urbanism*, chap. XXVIII in M. Foth (ed.) *Handbook of Research on Urban Informatics*. Hershey, PA, IGI Global, pp. 414-436.
- Perng, S.Y., Kitchin, R. and Evans, L. (2016) *Locative media and data-driven computing experiments*, in “Big Data and Society”, 3 (1), pp. 1-12
- Savransky, M. (2016) *The Adventure of Relevance: An Ethics of Social Inquiry*, London, Palgrave Macmillan.
- Star, S. L. (1999) *The ethnography of infrastructure*, in “American Behavioral Scientist”, 43 (3), pp. 377-391.
- Zuiderent-Jerak, T. and Jensen, C.B. (2007) *Editorial introduction: unpacking “intervention” in science and technology studies*, in “Science as Culture”, 16(3), pp. 227-235.

Open data re-use and data frictions. The tension between attachment, detachment and reattachment.

Antoine Courmont

Open data and smart cities policies underscore the implementation of the ideology of information liberalism into the urban government. This ideology was theorized by the French scholar Benjamin Loveluck through a political genealogy of Internet (Loveluck 2015). He claims that information is at the core of contemporary liberalism. Information must circulate freely in order to solve various problems in the context of cybernetic theories. The information liberalism is based on the assumption that data exist, are autonomous and can easily circulate.

Based on an ethnographic investigation inside the Metropolis of Lyon (France), I question this assumption. Indeed, the analysis of the open data policy *in the making* reveals a tension between attachment and detachment that needs to be addressed to allow a smooth circulation of data (Courmont 2016). Attached to vast socio-technical networks, data must be detached from their initial environment to circulate, before being re-attached to new users. Following traffic data from its production to its reuse (a perspective similar to a data journeys approach (Bates, Lin, and Goodale 2016), I will consecutively highlight the attachment between data and local transportation policies, the trials of detachment to make data circulate, and the data reattachment to secondary uses. The description of each of these steps illustrates how open data transform urban government.

Attachments: produce traffic data

Data are never autonomous neither immaterial. They are always entangled with and gifted by a collective of objects, people, techniques, ideologies, etc. Data are composed of various attachments (Gomart and Hennion 1999), which form a sociotechnical network made of heterogeneous entities. For instance, traffic data are part of a long chain from road sensors to traffic lights remote control in real-time by a specialized team inside a central control room. Algorithms, fiber optic network, data storage and even the traffic regulation policy cannot be separated from traffic data.

These attachments are never neutral: they do something. In this case, the traffic data produces a specific representation of the city that is used to regulate the road network. Indeed, data are based on conventions (Desrosières 2002), which define what must be represented to meet a specific use. The road network is not represented in its whole. Only some road sections are represented: the ones where the local authority want to tackle traffic congestion. That's why data are not only composed of attachments, but they also attach. They produce a link between a specific representation of the city and an actor who acts on it. In this case, traffic data attaches the

local transportation authority to its network through a convention defined to regulate road traffic.

Detachments: make traffic data circulate

To make data circulate, it is necessary to recompose the sociotechnical network of the data by untying some associations – the detachment – and constituting new ones – the reattachment. Indeed, the challenge is not to make data autonomous, but to ensure it can be well-attached to new users. The process of opening data is the result of a series of uncertain trials, during which the characteristics of the data, the producers and the users, are simultaneously re-defined. These *trials of diffusibility* recompose the sociotechnical networks of the data. To detach data from their initial environment, their previous ties are questioned in the light of their future attachment to prefigured users (Akrich 1992). This process changes the data. Moving from a trial to another one, data differs by the network it deploys. Data as stable and unchanging entity is a fiction.

While open data activists ask public bodies to release their data without thinking about the re-uses, in practice, prefigured users are constructed by the producers to decide to open or not their data, and, especially, how to open it. These usage scenarios vary depending on the producer or the data. A common fear is the risk of misuses or uses that may backfire on them. As instance, as Martin - a data producer - told me:

We do not want to cause any trouble to some projects defended by our colleagues. For instance, what if some people misuse historical traffic data to oppose infrastructure projects? Indeed, data may be used in the right way, but, these data are very technical, and it could also be quite difficult to interpret them correctly.

These prefigured uses determine the sensitiveness of the data. To overcome reluctances of the producers, the perceived risks are weighed by potential gains. Beneficial scenarios are also constructed and allow the detachment of the data from its initial environment and its attachment to new users. For instance, the open data project leaders often took as an example the case of a carrier using traffic data to optimize his delivery journey is often put forward, a re-use of open data aligned with the public policy.

Not only these prefigured uses determine if the data will be released, but they also affect the characteristics of the released data. Indeed, data are always shaped to meet a specific purpose and/or constrain certain uses. Data are transformed before their release to make possible the detachment from their production infrastructure and to facilitate their reattachment to new users. This process of “rawification” (Denis and Goëta 2015) of the data is the result of discussions with users in order to sustain the attachment with them. For instance, traffic data were initially published in the form of *traficolor* to ensure the coherence of information and to avoid misuses. However, after exchanging his views with some academic users, the

producer has decided to release a “rawer” data which now include occupancy and flow rates.

The release of data is the result of a simultaneous process of detachment of the data from its information infrastructure and its attachment to a new environment. Nevertheless, on the contrary of the processes of innovation analyzed by Goulet and Vinck (Goulet and Vinck 2012), the release of data does not imply the dissociation of all the ties linking the data and its initial environment. Firstly, because the data continue to be daily produced and used by public organizations. Secondly, because the attachment to new users cannot be successful if the data is fully detached from its initial environment: data will not be actualized, etc. That’s why the challenge of opening data is to achieve dealing with this tension between detachment and attachment.

Reattachment: re-use open data

In order to follow the chain of open data, we need to analyze their use by external actors. Far from enthusiastic hopes of economic development and democratic renewal, the first evaluations of open data policies noticed the relatively low uses of released data. A French open data advocates noted in a blog post in January 2013:

organizations are going through a period of doubts and depression: the data blues. [...] the multitude of technical, juridical, cultural and organizational challenges have left a bitter taste in the mouth of data re-users”⁶.

This reaction highlights the fact that an offer of data does not automatically meet a demand of data. The reuse of data raises coordination issues between heterogeneous social worlds. A lot of operations of cleaning, crossing, standardizing or articulating data are required to allow their attachment to a new information environment. While some mediations are removed, other are added, changing the socio-technical networks of the data. I would like to emphasize three politics of reattachment of data to new users: the consolidation, the homogenization and the articulation.

Consolidation

Produced to meet a specific use, the dissociation of the ties between the producer and the user of data endangers the solidity of open data. While it is impossible to be sure that the modalities of production meet the needs of the secondary users, data are threatened with deliquescence (Didier

⁶ Source: <https://libertic.wordpress.com/2013/09/24/vers-la-fin-du-baby-blues-de-lopen-data> (retrieved May 14, 2016)

2009). For instance, produced for the traffic regulation, the public database describing the road networks cannot be used to calculate routes as Jacques, a data re-user, explained:

These data are inoperable to make graph transversal, that is to say to move from a point to another one, because there are no structural nodes, which allow to say “From here to there, we can go, from there to here, we cannot.

To be used in another social world while preserving their solidity, data must become boundary-objects (Star and Griesemer 1989). The open data as boundary-object need to be both adaptable to be used in various context and robust enough to maintain a common representation of the urban space between the actors. Data have to obtain two crucial properties: an interpretive flexibility and a common infrastructure between these social worlds (Trompette and Vinck 2009). These two characteristics are not inherent to the data but depends on the situation where they are used.

Before all secondary uses, re-users realize two operations: “sourcing” and “cleaning” the data. The first one consists of the identification, the understanding and the estimation of the liability of the data, to be sure it will fill the re-user’s needs. The second one represents all the preliminary actions on the data before integrating it in a new informational environment. The two operations strengthen the consolidation of open data by estimating their interpretive flexibility, that is to say to check they can meet a secondary use without lose their initial meaning. However, they are not sufficient to allow the open data to become a boundary-object. A shared infrastructure between the producer and the re-user is required. Open data can easily cross organizational borders if it fills with the particular conventions of a professional field. If the traffic data are published in a standard format, it will be easily used by a traffic specialist and integrated in his own information system. For instance, this dataset was used by several companies specialized in traffic information: to publish real-time information in mobile apps or to make traffic prediction. As a result, open data as boundary-object allow the coordination of various actors through a common representation of the urban space defined by local authorities.

Homogenization

The homogenization is the production of a new aggregate from heterogeneous databases. The open data is a source of information among others to produce a new data which will standardize various representations of the urban space. For instance, Here Maps is a company providing mapping data to navigation services. The construction of these maps rests on various sources of information. Open data are used exclusively to update the initial database. To be associated to the Here’s database, open data must respect precise norms established by the company in order to ensure a high degree of quality. Realized by local workers, these operations of data qualification

and integration are largely invisible. However, they are crucial to smooth the numerous frictions which are inherent to the establishment of a relationship between informational infrastructures.

The construction of this homogenized database requires an equivalence convention in order to obtain a standardized representation of road networks all over the world. Unlike national statistical system, the definition of these conventions is no longer the sole privilege of the State or public actors. Private actors, like Here Maps, establish their own equivalence conventions, exposing public bodies to a loss of control over their public policies. Traffic regulation policies represent a good example of this risk. In this domain, public information services are in competition with private GPS services. The latter's road databases are based on an equivalence convention which differs from the hierarchy of roads defined by local transportation authorities. This difference of representation is not neutral: it is a prescriptive force to drivers through routes offered by GPS services. This is particularly apparent in case of congestion when the traffic is relocated to roads public authorities considered not suitable (minor road, etc.). The hierarchy of roads set by public authorities are no longer the convention, which reduces its ability to regulate the traffic regulation policy.

Articulation

The articulation is the third modality of open data reuse. It is characterized by the linking between various data through a common attribute. The heterogeneity of each data is preserved. For instance, the project Optimod, an intelligent transport system developed by the Great Authority of Lyon, whose aim is to gather, articulate and analyze data from all modes of transportation to offer multimodal information services. The differences of structuration of each data make impossible their homogenization in a common database. The challenge is to preserve the data inheritance by linking the datasets without change the way they are produced. An articulation work (Strauss 1988) is thus necessary using a common denominator. In the project Optimod, a geographic frame of reference, describing all the road network, was produced to allow the relationship between databases that were incompatible.

The outcome of this data articulation work is a new representation of the urban space. While transportation data represented the transportation network as a whole, data articulation offers a representation of all available transport modes according to the user's location.

From a representation of flow of vehicles to a specific representation for each traveler.

As a result, the target of policy moves from transportation networks to each individual. Transportation is not any more managed through the representation of flux of vehicles in a road network, but it is governed through individual travelers to which a singularized representation is offered. The articulation of data does not yield generalized representation, but it allows

particularization. Using articulated data, it is no longer the “we” which is governed, but each individual that becomes governable.

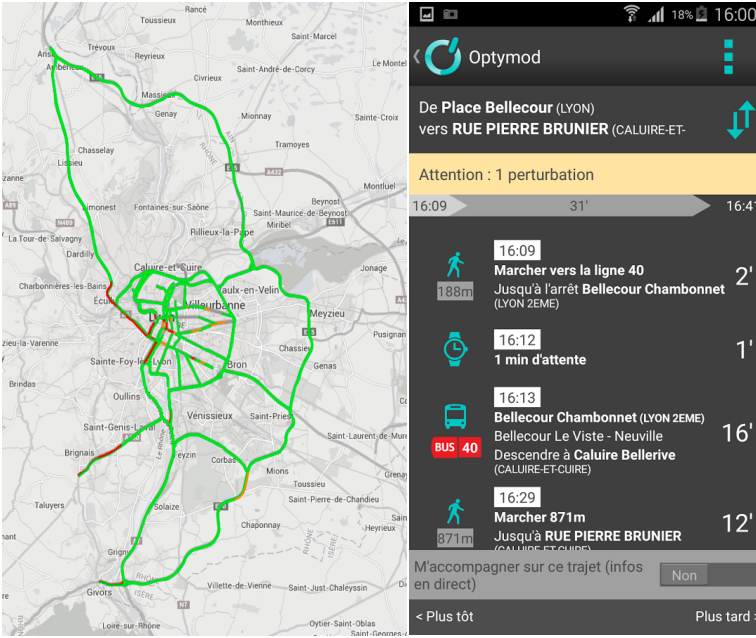


Figure 1 – From a representation of flow of vehicles to a specific representation for each traveler.

Conclusion

Drawing on empirical analysis of an open data policy, I have sought to contribute to the information infrastructure studies (Bowker et al. 2010; Edwards et al. 2009), by pointing out an inherent tension between attachment and detachment when making data circulate. Using the concepts of attachment and detachment to analyze the circulation of data highlights the sociotechnical network of a data and its necessary reconfiguration when data cross organizational borders. This aspect is crucial at the age of *big data* which place secondary uses of data at its core. Moreover, the attachment’s framework focuses on the attachments of data, but also on how the data itself attached (Gomart and Hennion 1999). Following the example of traffic data, I sought to underscore the joint redefinition of the data, the representation of the urban space and the institution acting on it. Moving from one social world to another one, the open data obtain new characteristics. This evolution of data changes the representation of the urban space, and, *in fine*, affects public policies.

- The consolidated data is a boundary-object allowing the coordination of various actors through a common representation of the urban space. The consolidation gives the producer new regulation opportunities by gathering these actors around his data.
- The homogenized data offers an alternative representation of urban space by the establishment of a new equivalence convention. The producer loses his control of the representation of the city which is a risk in case of conflict between these heterogeneous points of view.
- The articulated data gathers a diversity of points of view on a same object. In this way, it makes visible the singularity of each of these entities and makes possible their individual government.

This attachment/detachment framework opens new perspectives to analyze the reconfiguration of urban government in the age of information liberalism.

References

- Akrich, M. (1992) *The De-Description of Technical Objects*, in W.E. Bijker and J. Law (eds.) *Shaping Technology/Building Society Studies in Sociotechnical Change*, Cambridge, MIT Press, pp. 205-224.
- Bates, J., Lin Y. and Goodale P. (2016) *Data Journeys: Capturing the Socio-Material Constitution of Data Objects and Flows*, in "Big Data & Society", 3(2), pp. 1-12.
- Bowker, G.C., Baker K., Millerand F. and Ribes D. (2010) *Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment*, in J. Hunsinger, L. Kjastrup and M. Allen (eds.), *International Handbook of Internet Research*, New York, Springer, pp. 97-118.
- Courmont, A. (2016) *Politiques Des Données Urbaines. Ce Que L'open Data Fait Au Gouvernement Urbain*, Paris, Sciences Po.
- Denis, J., and Goëta S. (2015) *La Fabrique Des Données Brutes. Le Travail En Coulisses de L'open Data*, in C. Mabi, J.-C. Plantin and L. Monnoyer-Smith (eds.) *Penser L'écosystème Des Données. Les Enjeux Scientifiques et Politiques Des Données Numériques*, Paris, Editions de la Maison des Sciences de l'Homme.
- Desrosières, A. (2002) *The Politics of Large Numbers. A History of Statistical Reasoning*. Cambridge, Harvard University Press.
- Didier, E. (2009) *En Quoi Consiste l'Amérique? Les Statistiques, Le New Deal et La Démocratie*, Paris, La Découverte.
- Edwards, P., Jackson S., Bowker G.C. and Williams R. (2009) *Introduction: An Agenda for Infrastructure Studies*, in "Journal of the Association for Information Systems", 10(5), pp. 364-374.
- Gomart, E. and Hennon A. (1999) *A Sociology of Attachment: Music Lovers, Drug Addicts*, in J. Law and J. Hassard (eds.), *Actor Network Theory and After*, Oxford, Blackwell.

- Goulet, F. and Vinck D. (2012) *Innovation through Withdrawal. Contribution to a Sociology of Detachment*, in "Revue Française de Sociologie", 53 (2), pp. 117-146.
- Loveluck, B. (2015) *Réseaux, Libertés et Contrôle. Une Généalogie Politique D'internet*, Paris, Armand Colin.
- Star, S.L. and Griesemer J.R. (1989) *Institutional Ecology, 'Translations' and Boundary Objects: Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39*, in "Social Studies of Science", 19(3), pp. 387-420.
- Strauss, A. (1988) *The Articulation of Project Work: An Organizational Process*. in "The Sociological Quarterly", 29(2), pp. 163-178.
- Trompette, P. and Vinck D. (2009) *Retour Sur La Notion D'objet-Frontière* in "Revue D'anthropologie Des Connaissances", 3(1), pp. 5-27.

* * *

Transparency in the Rupture? Open Data and the Datafication of Society

Rolien Hoyng

According to the handbook by the non-profit organization Open Knowledge International, Open Data's definition centers on the availability of (digital) datasets "at no more than a reasonable reproduction cost," "in a convenient and modifiable form," and regardless of the fields of endeavor where they are applied.⁷ In this essay, I am interested in Open Data's coalescing with larger processes of datafication and the contradictions stemming from the combination of a discourse of transparency with the expedience of data for capitalism and algorithmic governance in the so-called smart city. That is to say, a profound ambiguity exists regarding what Open Data is all about. On the one hand, there is a promise for transparency, oversight, and mastery, building on Enlightenment epistemologies and notions of agency. Yet, on the other hand, we witness compounding datafication, namely the rendering into data of social processes and everyday life by means of (self)tracking in order to govern populations, markets, and cities (Cukier and Mayer-Schoenberger 2013). Despite its promotional discourse, Open Data does not reveal or open up a terrain of "pure" transparency and unmediated visibility. Instead, Open Data (re)produces regimes of visibility, enacting particular modes and distributions of perception and cognition (Birchall 2015; Flyverbom et al. 2016; Halpern 2014). In this essay, I address the implications of so-called 'smart', data-driven

⁷ See <http://opendatahandbook.org/guide/en/what-is-open-data/>

urbanism for citizenship.

Yet rather than associating the smart city's datafication with all-encompassing oversight and efficient control, I emphasize the messy and speculative character of data-driven applications, the productive role of errors and failure, and the elicitation of socio-technical emergence in smart urbanism. I argue that the critical import of 'transparency' needs to be reassessed in relation to these dynamics. If data activism should not be oriented onto 'opening up' the black-boxed smart city in order to restore transparency, what might it target instead?

Open Data draws from histories that invigorate the value of transparency and reinforce supposed relations between seeing, knowing, and acting. Bratich (2016, 178) raises the question of what the Enlightenment era would be "without a will to transparency?" He (178) continues: "[H]ow would modern communication persist without a similar desire—for openness, for clear channels, for a world without obscurity?" Transparency is connected to the dream of an entirely visible society, in which there is no "dark" corner remaining. Data and information play a key role in rendering society governable, but also governors accountable (Ananny and Crawford 2016, 2-3; Foucault 1979, 195-228). Accordingly, Open Data practices abide by a regime of visibility centered on "representation." Data function as evidence for what exists "out there" and possess a referential capacity (Halpern 2014, 46-51). Ordering data (through capturing, structuring, aggregating, and visualizing) forms part and parcel of ordering society as well as eradicating irrationalities, inefficiencies, and corruption. Especially the release of public service data carries with it the promise of rendering governance more efficient and holding governments accountable. For instance, Open Data initiatives can publish datasets on air or beach water quality that enable others to build apps or otherwise inform people about when to go out or swim. But with the data at hand, people could also find ways to track environmental quality over time in different areas in order to hold governors accountable for the state of affairs. Other datasets pertaining to government operations may assist in analyzing and visualizing politicians' voting behaviors or governments' budget expenditures.

Yet Open Data also supports datafication and the algorithmic governance of targeted populations and markets. As many critics of the smart city have argued, datafication is concomitant with the expansion of society's technological cognitive nonconscious (Hayles 2014), advancing covert forms of social sorting, profiling, modulation, and control to which populations are subjected (Deleuze 1992; Lyon 2001). Datafication draws from histories of cybernetics, which have reformulated cognition or intelligence in terms of rationality rather than reason. Cybernetics render human agency more or less intrinsic to preset computational rules, whereby small decisions are made in decentralized fashion that may be rational in the sense that they follow certain logics but that do not live up to the ideal of reason of a sovereign human subject standing apart from its environment (Halpern 2014, 173-191). When consumer-citizens in the smart city engage

with the interfaces of data-driven governance, they are often merely interpellated as data operators, who input data and act upon the feedback (Gabrys 2014). Interactivity is enabled as much as constrained by the design and protocols of the interface and often does not provide oversight of larger processes beyond self-tracking, let alone capacity for sovereign decision-making. As discussed in Young's contribution (in this issue), corporations are primary users of public datasets, as they combine Open Data with, for instance, social media data and data gained by tracking eye and body movements. Whereas this enables corporations to segment markets with increased granularity and predict consumers' behaviors and proclivities, whatever data their activities generate remain proprietary, as do the algorithmic applications deployed to process data. At stake is the paradox that Open Data despite its allusions to open access and transparency may be implicated in the advancement of digital enclosure (Andrejevic 2007) and empower corporate actors rather than citizens and the public at large.

Yet any account that portrays the datafied smart city only in terms of efficient control misses the following: data are often much less precise and all-encompassing than generally presumed, and algorithmic governance much more speculative, tentative, and prone to failure than expected. It may be the "messy" qualities of smart urbanism, rather than the ability to conclusively surveil and order, that deserve our analytical and critical attention. For instance, security tools in the smart city correlate heterogeneous datasets in order to calculate probability and risk. Yet risk-calculating derivatives do not draw conclusions on the basis of precise data with indisputable referential qualities. Instead, such derivatives infer and project on the basis of "uncertain and indifferent relationalities of missing elements" (Amoore 2011, 38). Rather than truthful representation, "[w]hat matters instead is the capacity to act in the face of uncertainty, to render data actionable" (Amoore 2011, 29). Not only are mistakes – false hits in the context of security – they also do not form systemic weaknesses. As long as mistakes provide feedback that helps the system evolve, they are productive: "The false hits of multiple security interventions that prove negative can never be errors in the terms of the derivative, for they too are folded back into association" (Amoore 2011, 32). Similarly, reviewing the development of the smart city of Songdo in Korea, Halpern et al. (2013) discuss test-bed urbanism as a way of experimenting with the management of urban space and life by means of extensive tracking. This extensive tracking does not imply the production of order through knowledge and surveillance. Instead, Halpern et al. (2013, 295) refer to smart city urbanism as a "new form of administration that lacks norms, frequency distributions, and the statistical apparatus of older demographic, state, and economic thinking in the name of a new epistemology of infinity, nonnormativity, and speculation." Smart urbanism operates through the uncertainties of speculation, trial, and emergence. Datafication here stands in the service of the production of value by means of innovation, for which the instance of failure is a driver rather than an obstacle, as long as its data result in the "next"

thing. In this context, constant testing and versioning replaces decisive conceptualization of failure and loss.

The messy character of algorithmic governance somehow corresponds to smart urbanism's organizational models. Whereas the tactical, speculative interventions of data-driven governance undermine disciplinary techniques of statecraft, state authority (at least to some degree) relinquishes centralized organization, overview, and control. Though such narratives still demand critical engagement, smart urbanism evokes scenarios according to which cognition is not centralized and rational (as is state authority); instead it takes its cues from swarms, insect colonies, and chaotic systems that inform models for self-organization (Halpern et al. 2013). In this present issue, Blok and Minor's discussion of governance practices such as 'living laboratories' and participatory design provides an empirical account of the extent to which smart city governance seeks to harness a degree of socio-technical contingency, rather than contain it, in order to multiply the effects of data across the smart-city environment. Accordingly, Thrift (2014, 6) proposes to see sentient cities as "spaces of ramification as different kinds of edge structure" and "as refuges that encourage experiment, tinkering and other adaptive practices" that offer "new ways to produce chaos out of order [...]."

If data tracking in the smart city has to do with tactical intervention and tapping socio-technical emergence in environments that collide failure and success, where does this leave the value of transparency? Should critical analysis and data activism revolve around regaining transparency vis-a-vis the black-boxed smart city? The role of data in the production of order on the basis of reason – in other words, data as part of the coupling of power/knowledge – is challenged. But so is the value of transparency in the service of accountability in that the latter might not be able to tackle the open-ended quality of smart systems and processes of socio-technical emergence that exploit failure and are enabled by incompleteness. Discussing algorithmic systems operative in areas of public governance ranging from transport to healthcare and policing, Ananny and Crawford (2016, 9-10) argue that even system designers themselves might not be able to provide a clear picture of complex and dynamically changing, adaptive systems. But more, the demand to "open up" the smart city's black boxes and see any systems does not yet account for less immediate and more complex socio-technical ramifications and (unintended) emergences.

Alternatively, engaged struggle could target the distribution of perception and cognition and the potential for socio-technical emergence itself, which is one way to interpret data activism and hacktivism. Coté (2014) has suggested that data activism might revolve around the dualism of data mobility and motility. The former refers to the contained movement of data that "primarily augments the profitable growth of the business of BSD [Big Social Data] and new forms of digital state surveillance" (123). Those actors however "loathe autonomous data motility," which "signals a possible route for the progressive becoming of a new data commons" (124, 140).

Accordingly, data activism includes scraping, rescuing, and “freeing” data. In such instances, data are not simply accessible and “open” but seized and motile in order to be processed in various ways. To the extent that activism preys on infrastructural “cracks” and systematic weaknesses in order to seize and “free” data, it is not smart urbanism’s success and failure (as in test-bed urbanism) that start to merge but resistance and failure.

To sum up, regardless of the fact that smart urbanism is stimulated by actual governments – utilizing the public service Open Data these governments provide – datafication unfolds at the expense of centralized state power and statecraft. It implies an increase of proprietary data relative to public data as a result of smart urbanism undertaken by the private sector, even though states and especially their security and intelligence units can demand access to this data, too (Taylor and Broeders 2015; Van Dijk 2014, 203). The ensuing technological cognitive nonconscious (Hayles 2014) is however messier and less controllable than often assumed. Smart urbanism challenges disciplinary modes of statecraft intended to produce order by means of transparent oversight, notwithstanding the fact that the very linkage between data, transparency, and order undergirds the promise of fortified governance efficiency and accountability in Open Data discourses. The encounter between Open Data and datafication as the de facto un-mappable expansion of society’s technological nonconscious generates contradictions and concocts a field of struggle. “Seizing” data and rendering it motile could form a tactic of resistance intended on releasing data in more radical ways than Open Data initiatives generally do. Critical questions are: what does it take for Open Data to become a site and medium for the expression of antagonistic struggles, in ways that belie Open Data’s semblance of neutrality rendered through claims to indiscriminately support all uses of open datasets? For instance, in what ways could tactical interventions exploit Open Data’s regimes of visibility in order to generate alternative modes and distributions of perception and cognition? And, how would such interventions radicalize the construction of “openness” in Open Data? But also, data activism itself requires more critical attention. Is it necessary to actually distinguish between failure and resistance and hence to further qualify those seizures and motilities that would be able to counter the disempowerment of variably-positioned bodies and different subjectivities in the smart city? Especially, if smart urbanism already is characterized by a degree of chaos and decentralization, in what ways could the seizures and motilities inflicted by data activism disrupt the ongoing disruptions incurred by smart urbanism?

References

- Amoore, L. (2011) *Data derivatives on the emergence of a security risk calculus for our times*, in “Theory, Culture & Society”, 28 (6), pp. 24-43.

- Ananny, M. and Crawford, K. (2016) *Seeing without knowing: limitations of the transparency ideal and its application to algorithmic accountability*, in "New Media and Society", Article first published online.
- Andrejevic, M. (2007) *Surveillance in the digital enclosure*, in "The Communication Review", 10 (4), pp. 295-317.
- Birchall, C. (2015) "*Data.gov-in-a-box*" *delimiting transparency*, in "European Journal of Social Theory", 18 (2), pp. 185-202.
- Bratich, J. (2016) *Occult(ing) transparency: an epilogue*, in "International Journal of Communication", 10, pp. 178-181.
- Coté, M. (2014) *Data motility: The materiality of big social data*, in "Cultural Studies Review", 20 (1), pp. 121-49.
- Cukier, K. and Mayer-Schoenberger, V. (2013) *The rise of big data: how it's changing the way we think about the world*, in "Foreign Affairs", May-June, pp. 28-40.
- Deleuze, G. (1992) *Postscript on the societies of control*, in "October" 59, Winter, pp. 3-7.
- Flyverbom, M., Leonardi, P.M., Stohl, C. and Stohl, M. (2016) *The management of visibilities in the digital age: introduction*, in "International Journal of Communication", 10, pp. 98-109.
- Foucault, M. (1979) *Discipline and Punish: The Birth of the Prison*, New York, Vintage.
- Gabrys, J. (2014) *Programming environments: environmentality and citizen sensing in the smart city*, in "Environment and Planning D: Society and Space", pp. 32 (1), pp. 30-48.
- Halpern, O. (2014) *Beautiful Data: a History of Vision and Reason since 1945*, Durham, Duke University Press.
- Halpern, O., LeCavalier, J., Calvillo, N. and Pietsch, W. (2013) *Test-bed urbanism*, in "Public Culture", 25(2), pp. 273-306.
- Hayles, N.K. (2014) *Cognition everywhere: the rise of the cognitive nonconscious and the costs of consciousness*, in "New Literary History", 45 (2), pp. 199-220.
- Lyon, D. (2001) *Facing the future: seeking ethics for everyday surveillance*, in "Ethics and Information Technology", 3, pp. 171-181.
- Taylor, L. and Broeders, D. (2015) *In the name of development: power, profit and the datafication of the global south*, in "Geoforum", 64, pp. 229-237.
- Thrift, N. (2014) *The 'sentient' city and what it may portend*, in "Big Data & Society", 1 (1), pp. 1-21.
- Van Dijck, J. (2014) *Datafication, dataism and dataveillance: big data between scientific paradigm and ideology*, "Surveillance & Society", 12 (2), pp. 197-208.

The ecologies of open data labor. A case study of the coproduction of an open geographical data base

Clément Marquet

This contribution concerns the collaboration between Transilien, a public transport operator on the region Île-de-France⁸ and the association Open Street Map (OSM) France. OSM is an open crowdsourced geographical data base, mainly produced by volunteers (Goodchild 2007). The partnership aims at asking the OSM volunteers to map the accessibility for the disabled equipments in 90 of the 380 railway stations composing Transilien's network. The mapping has to be done during the summer 2013, between June and September. Indeed, as part of Transilien open data policy, these geographical data has to be valued during a hackathon⁹ in November, in order to test how useful they can be for software developers.

However, in September 2013, Vincent, Transilien's chief of project discovers that only half of the mapping has been done. The volunteers did not respond to the call made by the association. With the spokesperson and the president of OSM France, Vincent has to contribute to the mapping, mainly during his free time. Though the mobilization of the volunteers is seen as a failure, the geographical data are widely used by the participants during the hackathon and Transilien decides to keep mapping its station on OSM. But the mapping can't rely anymore on the volunteers, Transilien considers paying people for that. As we will see, mapping is a tedious activity, and we will wonder how this tiresome and volunteered practice is turned into a paid work, within the open data ideology of "doing more with less".

Having to produce the data one wants to open is a typical situation in open data projects, as Jérôme Denis and Samuel Goëta have shown (Goëta 2016; Denis and Goëta 2017). Data rarely exist in the format desired and opening data necessitates a lot of work of identification, extraction, cleaning, etc. On the same trend, Antoine Courmont (2015) states that opening one's data implies reframing one's information infrastructure (Star and Ruhdeler 1996) to take into account cleaning processes and alternative uses of data by external actors.

Collaboration between Transilien and OSM implies producing an "open" data base in two dimensions: first, one that can be reused by software developers, second, one that can be completed by anyone (the only condition is to be registered in OSM). Wondering how the OSM data platform becomes a boundary-object (Star and Griesemer 1989) between OSM

⁸ Transilien is a subsidiary company of Société Nationale des Chemins de Fer (SNCF), the main railway transport company in France.

⁹ Hackathons consist of generally 48-hours software developers contests.

community and a public transport operator, we will pay a close attention to the organizational and labor transformations: what forms of data labor are invented by Transilien? What kind of competences are recognized in geographical mapping? How does those transformations relate to the OSM community?

I/ Mapping practice

To understand the lack of mobilization of the volunteers, it is important to have a clearer idea of what kind of an activity mapping is. OSM is generally presented as a platform where everyone, once registered, should be able to contribute. To take part in drawing the map, one does not need expensive tools: the basic process is to print the area to be mapped on “walking papers” (figure 2) and start listing the items lacking in the public space, whether those items would be buildings and roads, or traffic lights, trees and pedestrian crossings. To help her in her task, one could use a GPS or smartphone, to record geographical position and take pictures of the area.



Figure 2 – Walking paper. Source: http://wiki.openstreetmap.org/wiki/File:Surveying_with_walking_papers.jpg (CC-BY-SA 2.0+).

Once the outdoor mapping is done, the contributor has to turn it into data, that is “machine readable” information in the right format. To do so,

the volunteer has a various set of free software to use, more or less user friendly, according to his level of expertise. User has to draw dots, lines, polygons, and tag them using information resources already in the software or coordination tools available online.

Thus, mapping starts with outdoor activity, or indoor, concerning the stations. However, having to go in each station on the railway network is not an appealing perspective for the volunteers. Furthermore, the specific accessibility equipment is difficult to be seen for one who isn't used to seeing it. Last and not least, mapping is also presented as a meticulous and tedious activity. OSM volunteers describe themselves as “the ants” and Gael does not hesitate to qualify as “fucking boring” the moments when one “is counting steps under the rain” or adding data “until your hands hurt”. But, he adds, “at the end it's perfect, and that is what makes the beauty of it.”

2/ Paying the mappers: from dirty work to external expertise

Organizing the production of data in OSM is not a linear process for Transilien, but rather a trial and error process. From October 2014 to June 2016, Transilien uses three methods to produce data about its network.

A/ Mapping as a dirty work

To produce a large amount of data on the 290 stations left in a short period of time, Transilien hires students. Sixteen students are trained to OSM software and field mapping by OSM France leaders. They have roughly access to the same material and information as OSM contributors: specification note, walking papers, camera and smartphone GPS. The kind of contract linking the student to the firm is quite loose. As a precarious job, students are paid “at the station”. Some will never go through all the stations they should do, others will complain that the data they harvested have been changed or rejected by contributors. According to Vincent, this experience was tough because of the problems of skills and the lack of coordination.

Here, mapping appears to be a dirty work (Hughes 1962). If we recall the tedious process described by Gael, we understand that this process is delegated by Transilien to students. As they do not see the beauty of it and as the added value to their formation is quite slim, recognition is hard to be found. Their coordinator clearly states that this is far from being its main project at the time.

B/ Internalizing competences through an OSM contributor

In 2015, a change is made in both the managing and the strategy of the mapping. For personal reasons, Vincent has to stop working in Transilien,

and he is temporarily replaced by an OSM contributor, Florian. He explains that the work of the students was too messy: the tags weren't coherent enough, the mapping was not accurate. He doesn't blame them, as they were nearly starting from blank slate. However, most of their work is finally erased and Florian reorganizes the mapping procedure. Hence, he draws map from his office, using architects' plans, rather than going on the field. Once he has drawn the 379 stations, he goes checking what has changed, what could be added or deleted. However, despite of his efforts, 8 stations resist to the mapping. The most complex, multilevel stations are far too difficult to be mapped from the only basis of architect plans, even for an experienced mapper.

C/ High expertise mapmaking: OSM contributors as service providers

Through Florian, Transilien contacts a geomatics company named Cartocité. This agency is composed by three regular OSM contributors specialized in hardware and software development, and in geomatics. It is the first time the agency signs a contract to realize OSM mapping. Such a structure was necessary in particular because of confidentiality contracts. Indeed, mapping the stations gives the agency access to the architectural layouts of the station, as non public areas should not be mapped in OSM. When interviewed Antoine sees this as a privilege, in his point of view, many OSM mappers would have been delighted to be authorized to map a station through architects layouts, without even being paid for that. Once they have the layouts, Cartocité employees enter in a three months process, involving software and hardware development (a four camera pod to take 360° pictures), plus the advanced knowledge of geomatics and OSM software.

The first steps of the production of railway stations open data reminds classic information gathering across history, from the birth of statistic polling, which consisted in hiring unemployed people (Didier 2009), to recent practices of inquiries which also rely on low qualified jobs (Caveng 2012). Data harvesting is considered by Transilien as a dirty work (Hughes 1962), that could be done with little investment. However, contrary to opinion polls, the geographic data expected by Transilien needs high accuracy. Furthermore, mapping is not considered as a dirty work for everybody. As a relational notion, what is dirty and boring to some is not for others: most of OSM contributors see mapping activity as a – tedious – leisure (Dufael et al. 2016).

Thus, Transilien takes into account complexity of data production as an issue of competences and labor perception, and hires OSM contributors and experts accordingly. However, by doing so, the leisure becomes a labor, which questions Cartocité' employees. This episode gives us an insight regarding the tensions coming with open data transformations: administrations tend to expect producing value with low investments, considering data as a simple asset, with few considerations of the work that must be

done to make it “open”. Taking this work into consideration needs, as we will see, to rethink the ecology of data labor.

3/ Maintaining the data: coordinating experts, agents and the crowd.

However, having turned stations into data isn't enough to have an open data platform. Opening data aims at the circulation of data, which means that many operations are required to let the data flow (Courmont this issue). Data have to be updated to follow up the transformation of the physical spaces. Other forms of data labor and organizational transformations are needed. To deal with data maintenance, Transilien divides the issue in two dimensions, both reframing data labor: updating and surveillance.

A/ An app to update the map

To have data up to date, and considering the failure of relying on the only OSM regular contributors to do so, Florian have the idea to develop a smartphone app. The aim of the app is to simplify the contribution process on the station. Anyone could add data in a few seconds, avoiding the more or less user-friendly editing software that are repulsive to the newcomers. In June 2017, a prototype is delivered. According to the management, it should be used by the travelers but also by the agents.

Thus, by doing this application, Transilien and OSM contributor Florian are opening another potential transformation in the OSM model of data production. By simplifying the addition of data, the app could move the model of production from a community level, in which volunteered are getting more and more entangled with professionals, to a crowd model, which allows a widening of the contributors but also lightens the links between them (Haythornwaite 2009).

B/ Surveilling data

With the open data base, the company faces new risks such as malevolence, errors or “tag wars”, which corresponds to disagreement between contributors regarding how to tag an item (Mooney 2011). The company could not accept this kind of volatility if services rely on the open data base. To overcome this situation, Transilien hires Cartocité. The agency has to develop software to monitor data activity on the stations. This makes the agency accountable enterprise for the station's data and stabilizes the role of the company as an “OSM professional”.

Many contributors developed tools in order have a watch on specific parts or items on the map and insure quality of data (Goodchild and Li 2012). As it was the case with the intensification of data production on complex station in the first part, we assist at a “formalization and an industrialization of practices that already existed” - but not at as a paid service.

C/ New roles for the agents

At the intersection of both the surveillance and the updating program, we can find transformations in the mundane work of Transilien agents. Indeed, Cartocité's surveillance process is to send the dubious data to Transilien employee who then send the data to the head of the station, who asks the agents to check whether the modified items are corresponding to the reality or not. Furthermore, the updating app, Mapmagare, has a version dedicated to agents. Thus, either checking what has been modified, or registering fresh transformations, agents become in charge of the stations data.

This second moment in the organization of the open data infrastructure let us see the production of an ecology of data maintenance articulating external companies, Transilien managers agents and a hypothetical crowd. The issue is not anymore about the harvesting of data but concerns the management of the "data lifecycle" across time, which demands more coordination, mostly assumed by the production of new software: monitoring software, for Cartocité, mapping software, for the agents and the crowd.

The transformation in Transilien organization also have effects in the OSM community, as the OSM France association is now wondering how to deal with the professionalization of the contributors. Should the association get paid workers to be able to respond to demands such as Transilien's one? In June 2017, the bureau of OSM France have decided that its role would be to give visibility on their website to the self declared "OSM professional", but that they will not recommend one or another not to create inequalities amongst them.

Conclusions

By focusing on the collaboration between various kinds of actors to produce a geographical data set in an open data base, this commentary gives an insight of the situated inventions of open data labor and of the organizational transformation that goes with this invention.

Though the open data platform is shared between Transilien and OSM, Transilien is taking the most active part in the production and the maintenance of the data set. The direct contribution of business practices in the open data base could have been seen as a problem, like it is in many open source communities (Demaziere et al. 2009). According to the various actors implied in the process (Cartocité, OSM France association, Transilien, and a few OSM contributors), the OSM basic rules of contribution and data license provide a framework to allow the entanglements of public and private interest in the data production and maintenance.

Along this common goal we can see many attempts from Transilien to fine a suited coordination of actors to map its station. Indeed, though Transilien accepts to play by OSM rules in the production of data, the company also tried to make this production as cheap as possible. Paying a close attention to the production and the maintenance of the data reveals the variety of the forms of data labor experimented in the collaboration along with the coordination of various models of data maintenance, articulating OSM community with a wider crowd, agents mundane work and a surveillance company.

Finally, we can see how open data platforms and the original public private partnerships made around them (Young this issue) contribute to blur the boundaries of what counts as work (Strauss and Star 1999) in the liberal information society.

References

- Courmont, A. (2015) *Open data et recomposition du gouvernement urbain: de la donnée comme instrument à la donnée comme enjeu politique*, in “Informations sociale”, 191(5), pp. 45-50.
- Demaziere, D., Horne, F. and Zune, M. (2009) *Les développeurs de logiciels libres: militants, bénévoles, professionnels?* in D. Demazière and C. Gadéa (eds.), *Sociologie des groupes professionnels*, Paris, La Découverte, pp. 285-295.
- Denis, J. and Goëta, S. (2017) *Rawification and the careful generation of open government data*, in “Social Studies of Science”, 30 (5), pp. 604-629
- Didier, E. (2009) *En quoi consiste l'Amérique? les statistiques, le New Deal et la démocratie*. Paris, La Découverte.
- Duféal, M., Jonchères, C. and Noucher, M., (2016) *Ecce carto – Des espaces de la contribution à la contribution sur l'espace - Profils, pratiques et valeurs d'engagement des contributeurs d'OpenStreetMap (OSM)*, UMR 5319.
- Goëta, S. (2016) *Instaurer des données, instaurer des publics Une enquête sociologique dans les coulisses de l'open data*, PhD Dissertation, Télécom ParisTech.
- Goodchild, M. (2007) *Citizens as sensors: The world of volunteered geography*, in “GeoJournal”, 69, pp. 211-221.
- Goodchild, M. and Linna L. (2012) *Assuring the quality of volunteered geographic information*, in “Spatial statistics” 1, pp. 110-120.
- Haythornthwaite, C. (2009) *Crowds and communities: Light and heavyweight models of peer production*, in “System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference IEEE”, pp. 1-10.
- Hughes E. (1962) *Good People and Dirty Work*, in “Social Problems”, 10(1), 3-11.
- Lin, W. (2013) *Volunteered geographic information and networked publics? Politics of everyday mapping and spatial narratives*, in “GeoJournal” 78(6), pp. 949-965.

- Caveng R. (2012) *La production des enquêtes quantitatives*, in “Revue d'anthropologie des connaissances” 6(1), 65-88.
- Mooney, P. (2011) *The evolution and spatial volatility of VGI in OpenStreetMap*, in “Hengstberger Symposium Towards Digital Earth: 3D Spatial Data Infrastructures”, Heidelberg, September, pp. 7-8.
- Star, S.L., and Griesemer, J. R. (1989) *Institutional ecology, translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39*, in “Social studies of science”, 19(3), pp. 387-420.
- Star S.L. and Ruhleder, K. (1996) *Steps toward an ecology of infrastructure: Design and access for large information spaces*, in “Information systems research”, 7(1), pp. 111-134.
- Star, S. L., & Strauss, A. (1999) *Layers of silence, arenas of voice: The ecology of visible and invisible work*, in “Computer supported cooperative work (CSCW)”, 8(1), pp. 9-30.

* * *

Neo-Environmental Sensing: Ontological approaches to public data

Christian Nold

Introduction

This text offers a preliminary scoping of what I call ‘neo-environmental’ sensing. In the last decade there has been a radical change in environmental sensing, with hardware becoming cheaper and involving the public in data gathering. ‘Neo-environmental’ sensing takes place outside of governmentally mandated monitoring in the context of ‘participatory sensing’, ‘citizen science’ and ‘smart cities’ and uses networked technologies. The most commonly cited example is the Safecast radiation monitoring network that emerged in response to the Fukushima nuclear disaster (Safecast 2011). The hardware built by volunteers provided vital data for the public while the governmental response was criticised. Since this incident, there has been an enormous growth in low-cost environmental sensor systems built by hobbyists, entrepreneurs and research projects. These sensing devices are often crowdfunded via platforms such as Kickstarter and Indiegogo with the intention that people install them in their homes or carry them with them every day. The term ‘neo-environmental’ sensing is a reference to the ‘neogeography’ movement (Turner 2006) that emerged in 2006 and

was often attributed to the newfound public access to global positioning systems (GPS) and Web 2.0 technologies. Neogeography brought a new range of people to develop and use geographical mapping systems in ways that did not follow established protocols: “rather than making claims on scientific standards, methodologies of Neogeography tend towards the intuitive, expressive, personal, absurd, and/or/artistic, but may just be idiosyncratic applications of ‘real’ geographic techniques” (Eisnor 2006). The fact that neogeography did things ‘differently’ led to a range of epistemic and ontological tensions about how mapping practices can create truths (Warf and Sui 2010). In practice this often emerges as conflicts around data quality and power dynamics with volunteers, as illustrated in the text by Marquet (this issue). My paper involves a preliminary sketch of ‘neo-environmental’ sensing to identify how it differs from existing environmental monitoring that is based on specialised, calibrated sensor hardware. Some observers are already starting to question the impact of this new sensing paradigm (Kumar et al. 2015). My paper suggests ‘neo-environmental’ sensing presents a challenge to pollution-affected communities but also offers potential for ontological translation and contestation. In addition it creates a new role for academic researchers to help communities translate ‘smart’ data into matters of concern. The paper is based on my PhD research where I analysed four ‘neo-environmental’ devices over a period of years from design, usage and output (Nold 2017).

Environmental sensing via publicity

Institutional sensing of air and noise pollution involves large stationary hardware that costs tens of thousands of euros and is focused on long-term trends and regulatory standards. Another class of portable devices costs thousands of euros, and is used in response to specific pollution incidents. The cost and complexity of this hardware puts them out of the reach of pollution-affected communities. However, ‘neo-environmental’ sensing devices use ‘free’ inbuilt smartphone sensors or hardware that only costs hundreds of euros. The tradeoff for this accessibility is their limited capabilities that cannot differentiate pollutants, are often uncalibrated and are affected by temperature and humidity. Yet, ‘neo-environmental’ sensing devices offer very sophisticated networking capabilities, data repositories and APIs. The focus is not on the individual measurement instrument but on creating large scale sensing networks and visualisations. These visualisations are often real-time and graphically more sophisticated than governmental webservices. The goal is quantity and interoperability of data following concepts such as ‘smart cities’ (Batty 2012) and an ‘internet of things’ (Ashton 2009). Crucially, ‘neo-environmental’ sensing devices tend to be accompanied by a vast range of buzzwords and publicity. The common narrative is that sensing devices are ‘smart’ technologies that bring disruptive potential. A frequent claim is that there are more mobile phones than people on earth (Alfonso et al. 2015) and that this will create a global

sensing network where the

planet earth will don an electronic skin. [...] It consists of millions of embedded electronic pollution detectors, cameras, microphones [...] These will probe and monitor cities and endangered species, the atmosphere, our ships (Gross 1999, par. 2).

The argument is that digital networked sensing will lead to new ecological management as illustrated by articles such as *How Two Billion Smartphone Users Can Save Species!* (Preece 2017). Smart technologies allow the cost and labour of environmental sensing to be passed onto the public (Silvertown 2009). Some suggest these networked sensing platforms will become alternatives to governmental institutions (Townsend et al. 2010) and generate new technological citizenship (Kera et al. 2013). Kresin (2013, par. 3.) argues,

we know how to measure ourselves and our environment, to visualise and analyse the data, to come to conclusions and take action. [...] We are ready. But, as yet, our government is not.

Other researchers suggest that gathering environmental data will make participants more supportive of technological and scientific progress (Bonney et al. 2009) and shape environmentally beneficial behaviour (Maisonneuve et al. 2010; EveryAware 2011). In this framing, environmental sensing devices are no longer just sensors of external pollutants but become persuasion actuators that attempt to transform the user of the sensing device. In contrast, older, analog sensing devices such as diffusion tubes are hardly mentioned within the participatory sensing literature. Diffusion tubes have been used in the UK since 1976 (AEA Energy and Environment 2008) and consist of small plastic containers coated with chemical reagent that after exposure are sent to a certified laboratory. They are cheap and accurate ways of measuring air pollution, yet are not part of ‘neo-environmental’ sensing, since they do not contribute to global digital networks and mediagenic publicity.

These narratives highlight that ‘neo-environmental’ sensing is less concerned with material pollutants and health impacts and instead re-articulates the environment as data networks and mass involvement. I suggest that ‘neo-environmental’ sensing should be seen as more ‘expansive’ than traditional environmental monitoring. It involves a range of ‘big words’ (Bos et al. 2014) and ‘buzzwords’ (Bensaude-Vincent 2014, 250) that function to “create peaceful collectives of people with competing agendas. They act as a soft power attracting and enrolling people, thus preventing violence”. While these narratives have been successful in bringing together EU policymakers, academia, commercial entities and hobbyists, others such as pollution-affected communities have not been part of these narra-

tives. What is missing in the literature, and which this paper tries to remedy, is an acknowledgment that ‘neo-environmental’ sensing represents a challenge to existing public data practices.

Devices that shift subject and object

This paper provides a brief snapshot of how four ‘neo-environmental’ sensing devices impacted participants and engaged communities. In general, there was an ambiguity of what exactly the devices were sensing. Remarkably, all the devices started off sensing one entity and then shifted towards other phenomena. One device started off promising to sense pollution and offered radical political change, while at other times it was framed merely as a community of concerned people. Another device gave little detail about its hardware sensors but claimed to transform the public into smart citizens; while yet another device abandoned air pollution to focus on measuring the user’s mental awareness and behaviour. Often, users themselves became framed as the main subject of sensing, rather than external pollutants. Crucially, the ambiguity of what was being sensed existed also at a material level, where the devices were poor at differentiating phenomena and had sensors added over their lifetime. Two devices had interface sliders added to monitor and measure the behaviour of the users. Pollution data was often presented as raw values that users could not compare to official datasets. The devices often left the participants confused and frustrated and health impacts could not be meaningfully discussed. Nevertheless, the devices were all well funded, attracted many participants and were cited as good-practice exemplars within academic literature, EU policy reports and the mainstream media.

How was this possible? My suggestion is that ‘neo-environmental’ sensing does not function as an epistemic knowledge practice. The classic model of environmental sensing is premised on what Latour calls scientific ‘chains of reference’ (Latour 1999). These allow the backward tracing from a scientific report to the dataset and finally to the pollutant phenomena in the world. In this chain, sensing devices are meant to act simply as ‘intermediary objects’ that allow the progressive abstraction of the world into a scientific text or institutional report. Yet in ‘neo-environmental’ sensing, devices are not intermediaries and do not offer chains of reference. Instead they function as assemblages that combine a variety of different agendas. I suggest we should think of them as “patterned teleological arrangements” (Law and Ruppert 2013) to highlight the way they act as concentrations of agendas. This conceptualisation allows us to see how the devices fused together hardware with layers of rhetoric, visualisations and participants. By being ambiguous about ‘what was being sensed’, the devices could be detached from sensing material things such as pollution gases or sound vibration to become something more expansive and expressive. The device organisers often described the devices as ‘beacons’, ‘nodes’, ‘bridges’ and ‘ve-

hicles' towards something else. This ambiguity allowed the devices to articulate buzzwords of networked environments encompassing radicalism, smart citizenship and behaviour change. The result is that 'neo-environmental' sensing became much 'bigger' and more 'innovative' than traditional environmental sensing and able to generate more publicity and enrol more participants. In the framing as public engagement and digital network construction, 'neo-environmental' sensing devices became successful and won international awards even if they created a radically altered relationship towards the environment.

Translating 'smart' data into matters of concern

Yet this 'neo-environmental' approach caused problems for participants who had active health concerns or lived in pollutant-affected areas. Many of the participants were not aware of the ambiguous nature of the sensing devices and were often confused and frustrated and did not know what to do with the generated data. Traditionally political actors outside of science have often appropriated the credentialed 'intermediary objects' of science for their own purposes to create their own chains of reference in order to legitimate their environmental concerns. A well-documented example is the bucket brigades (Overdeest and Mayer 2007) that used containers to collect air samples, to be sent to a certified laboratory for analysis. In this way, legitimacy is embedded within the scientific instrument that is then 'borrowed' by a pressure group to make its localised argument. Kulenberg (2015, 67) suggests,

by turning to scientific methods in their political struggles, citizen scientists are able to 'short-circuit' the conventional modes of seeking political representation and use reference as a mediator in re-presenting the state of affairs that have come under controversy.

Similarly, Carton and Ache (2017) describe the potential of low-cost environmental sensing as opening a dialog with governments to strengthening the negotiating position of communities as 'information power'. Yet in my studies of organised deployments with 'neo-environmental' devices, groups that tried to use an epistemic logic of 'information power' could not make use of the data generated and led to the removal of the data from existing datasets. I argue, that 'neo-environmental' sensing does not support the borrowing of epistemic legitimacy, but instead requires a fundamentally different - ontological approach to environmental sensing. In this approach, sensing devices are used to deliberately enact multiple 'realities' as in Annemarie Mol's notion of ontology, where

ontology is not given in the order of things, but that, instead, ontologies are brought into being, sustained, or allowed to wither away in common, day-to-day, socio-material practices (Mol 2002, 6).

This ontological approach is best illustrated via a small case study of a ‘neo-environmental’ sensing device called WideNoise (EveryAware 2012). This smartphone noise-sensing app was originally created by a company as a technical demonstration of smart cities and internet of things and then later used by an academic research consortium. Crucially the app was uncalibrated and produced poor sound level measurements. Nevertheless, it proved to be surprisingly useful when used in relation to the contested issue of Heathrow airport expansion in London. A number of local actors managed to ontologically reconfigure the app to ‘sense’ a variety of entities that were relevant for the local controversy. Heathrow airport is the world’s third largest, and there are plans to expand it with an additional runway that will dramatically increase local air and noise pollution. The issue is a clash between different realities of environmental pollution as articulated and practiced by local residents, pressure groups and the commercial airport and politicians. In the context of a noise monitoring campaign jointly coordinated by a university and a pressure group, local residents and a council managed to re-purpose the app to enact new environmental ontologies and create connections towards institutional decision making on the runway expansion. The residents used the app to selectively measure the loudest planes that they found the most annoying. Their goal was to find an alternative to the current noise metrics that statistically average the measurement of noise events and thus underrepresent the sensorial shock of quick and loud over-flights. Crucially, the participants did not see the selective measurement of loud flights as manipulation of data but as adopting a rigorous experiential protocol. This could be clearly seen in the way the participants took care to be selective and avoid measuring non-aircraft noise. The residents were not trying to create exaggerated ‘fake’ data but were highlighting the ‘real’ high measurements that were occurring but being swamped by the averaging of the official noise metric. By selectively submitting noise data from planes that annoyed them, they were using the app to include their sensation within the regulatory ‘reality’ of noise that they felt excluded from. In a similar way, the local pressure group focused on the quantity of participants taking part in the noise monitoring campaign, yet largely ignored the decibel data. The monitoring campaign generated significant media publicity and the pressure group emphasised the act of public measurement as a mass protest against airport expansion. What mattered politically was the performative act of measurement and the quantity of participants rather than the epistemic content of the data. Finally, a local council made use of the generated data as the basis of their official response to the government’s consultative document on the third runway. They also did not focus on the decibel content of the dataset but highlighted qualitative textual descriptors used by residents and reframed the measurements as official noise complaints. Thus the existence and size of the dataset became evidence for the failure of the current institutional noise metrics to account for the experience of residents. In this way, the

app became a prototype for the new kinds of sensing devices that would be needed to better represent the residents' sensation of aircraft noise.

What these enactments of the app had in common was that they did not rely on epistemic chains of reference towards calibrated reference sensors or claims to the authority of science for their legitimacy. By politicising the supposed neutrality of the existing noise metrics and providing an alternative approach, these concern-based enactments managed to bypass reference devices as gatekeepers to environmental decision-making. The local actors validated their ontologies of noise by providing evidence of the strength of their concern via numbers of participants, intensity of complaints and media coverage of the monitoring campaign. The resulting legitimacy of these ontological translations was strong enough for the council to base their official response to the airports commission on the WideNoise app. In this case, the ambiguity of what the 'neo-environmental' device was sensing, offered potential for the local actors to translate the device, the act of measurement and the data into 'matters of concern' (Latour 2004). The case study demonstrates that while 'neo-environmental' sensing can create problems for pollution-affected communities when using an epistemic 'information power' approach, it can also allow the construction of new environmental realities.

Supporting ontological reconfiguration

My gut feeling is that 'neo-environmentalism' is here to stay and will continue to grow across a variety of different domains, as environmental politics becomes more technology and publicity driven. Yet rather than try to oppose this trend to return to naturalistic visions of stable epistemic data, my suggestion is that the trend presents researchers with an opportunity to shift towards a new approach and role. STS has long focused on analysing environmental controversies as epistemic conflicts of expertise and knowledge politics (Wynne 1992; Yearly 2000) and advocated on behalf of communities to articulate their knowledge claims. Yet arguably this approach is less useful in situations where environmental controversies revolve around ontological conflicts. I suggest that, in those cases, researchers should engage with the disruptive potential of 'neo-environmentalism' to redirect it towards multiplying realities as ontological politics (Mol 1999). Because 'neo-environmental' devices don't offer epistemic certainty, the devices invite a critical approach to controversies that politicises the way institutional standards function ontologically to exclude the realities of pollution-affected communities. I see a lot of potential in Marres (2013, 12) suggestion that 'ontology must be experimentalised', and that researchers should work with

the deliberate investment of non-humans with moral and political capacities. Here objects, and by extension ontologies, have political and moral capacities 'by design'.

This chimes with the way participatory designers use the notion of ‘infrastructuring’ (Björgvinsson et al. 2012; Dantec 2012) to embed the designer within a controversy and a community to support them over an extended period of time. My suggestion is that like neogeography, ‘neo-environmental’ sensing could become a movement for doing the environment ‘differently’, by working with local groups to support them in carrying out ontological translations. In the WideNoise study, the collaboration between the university researchers and the pressure group was critical for staging the sensing device in such a way that the multiple environmental enactments could take place. By focusing on the translation of ‘smart’ data into matters of concern, academic researchers could shift into a role of supporting communities in constructing new environmental ontologies.

Acknowledgments

This research has been supported by the European Commission via the DITOs project under the grant agreement No. 709443 as well as EU RD contract IST-265432 and the EPSRC Institutional Sponsorship awards to UCL 2011-12 (Grant References EP/J501396/1 & EP/ K503459/1).

References

- AEA Energy and Environment (2008) *Diffusion tubes for ambient NO₂ Monitoring: practical guidance for laboratories and users*, Technical report, Defra.
- Alfonso, L., Chacón, J. C. and Peña, G. (2015) *Allowing citizens to effortlessly become rainfall sensors*, in A. Mynett (ed.) *E-proceedings of the 36th IAHR World Congress*, Madrid, IAHR, pp. 1-5.
- Ashton, B. K. (2009) *That ‘Internet of Things’ Thing. In the Real World, Things Matter More than Ideas*, in <http://www.rfidjournal.com/article/view/4986> (retrieved July 20, 2017).
- Batty, M. (2012) *Smart cities, big data*, in “Environment and Planning B: Planning and Design”, 39 (2), pp. 191-193.
- Bensaude-Vincent, B. (2014) *The politics of buzzwords at the interface of technology, market and society: the case of ‘public engagement in science*, in “Public Understanding of Science”, 23 (3), pp. 238-253.
- Björgvinsson, E., Ehn, P. and Hillgren, P.-A. (2012) *Agonistic participatory design: working with marginalised social movements*, in “CoDesign”, 8 (2-3), pp. 127-144.
- Bonney, R., Cooper, C.B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K.V. and Shirk, J. (2009) *Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy*, in “Bioscience”, 59 (11), pp. 977-984.
- Bos, C., Walhout, B., Peine, A. and van Lente, H. (2014) *Steering with big words: articulating ideographs in research programs*, in “Journal of Responsible Innovation”, 1 (2), pp. 151-170.

- Carton, L. and Ache, P. (2017) *Citizen-sensor-networks to confront government decision-makers: two lessons from the Netherlands*, in "Journal of Environmental. Management", 196, pp. 234-251.
- Dantec, C. A. L. (2012) *Participation and publics: supporting community engagement*, in "Proceedings of the SIGCHI Conference on Human Factors in Computing Systems", ACM, pp. 1351-1360.
- Eisnor, D.-A. (2006) *What is neogeography anyway?*, in http://platial.typepad.com/news/2006/05/what_is_neogeog.html (retrieved July 20, 2017).
- EveryAware (2011) *EveryAware / Enhance Environmental Awareness through Social Information Technologies*, in <http://www.everyaware.eu> (retrieved July 20, 2017).
- EveryAware (2012) *WideNoise*, in <http://cs.everyaware.eu/event/widenoise> (retrieved July 20, 2017).
- Gross, N. (1999) *The Earth will don an electronic skin*, in http://www.businessweek/1999/99_35/b3644024.htm (retrieved March 27, 2013).
- Kera, D., Rod, J. and Petrova, R. (2013) *Post-apocalyptic citizenship and humanitarian hardware*, in R. Hindmarsh (ed.), *Nuclear Disaster at Fukushima Daiichi: Social, Political and Environmental Issues*, New York, Routledge, pp. 97-115.
- Kresin, F. (2013) *A Manifesto for Smart Citizens*, in <http://waag.org/en/blog/manifesto-smart-citizens> (retrieved July 20, 2017).
- Kullenberg, C. (2015) *Citizen science as resistance: crossing the boundary between reference and representation*, in "Journal of Resistance Studies", 1 (1), pp. 50-76.
- Kumar, P., Morawska, L., Martani, C., Biskos, G., Neophytou, M., Di Sabatino, S., Bell, M., Norford, L. and Britter, R. (2015) *The rise of low-cost sensing for managing air pollution in cities*, in "Environment International", 75 (February), pp. 199-205.
- Latour, B. (1999) *Circulating reference: sampling the soil in the amazon forest*, in *Pandora's Hope*, Cambridge MA and London, Harvard University Press, pp. 24-79.
- Latour, B. (2004) *Why has critique run out of steam? From matters of fact to matters of concern*, in "Critical Inquiry", 30 (2), pp. 225-248.
- Law, J. and Ruppert, E. (2013) *The social life of methods: devices*, in "Journal of Cultural Economy", 6 (3), pp. 229-240.
- Maisonneuve, N., Stevens, M. and Ochab, B. (2010) *Participatory noise pollution monitoring using mobile phones*, in "Information Polity", 15 (1-2), pp. 51-71.
- Marres, N. (2013) *Why political ontology must be experimentalized: on eco-show homes as devices of participation*, in "Social Studies of Science", 43 (3), pp. 417-443.
- Mol, A. (1999) *Ontological politics. a word and some questions*, in J. Law and J. Hassard (eds.), *Actor Network Theory and After*, Oxford and Malden, Blackwell Publishers, pp. 74-89.

- Mol, A. (2002) *The Body Multiple: Ontology in Medical Practice*, Durham and London, Duke University Press.
- Nold, C. (2017) *Device Studies of Participatory Sensing: Ontological Politics and Design Interventions*, PhD thesis, UCL, London.
- Overdeest, C. and Mayer, B. (2007) *Harnessing the power of information through community monitoring: insights from social science*, in “Texas Law Review”, 86, pp. 1493-1526.
- Preece, J. (2017) *How two billion smartphone users can save species!*, in “Interactions”, 24 (2), pp. 26-33.
- Safecast (2011) *Safecast: Open Environmental Data for Everyone*, in <http://blog.safecast.org> (retrieved July 20, 2017).
- Silvertown, J. (2009) *A new dawn for citizen science*, in “Trends in Ecology & Evolution”, 24 (9), pp. 467-471.
- Townsend, A., Maguire, R., Liebhold, M. and Crawford, M. (2010) *A planet of civic laboratories: the future of cities, information, and inclusion*, Technical report, Palo Alto, Institute for the Future.
- Turner, A.J. (2006) *Introduction to Neogeography*, Sebastopol, O’Reilly Media.
- Warf, B. and Sui, D. (2010) *From GIS to neogeography: ontological implications and theories of truth*, in “Annals of GIS”, 16 (4), pp. 197-209.
- Wynne, B. (1992) *Misunderstood misunderstanding: social identities and public uptake of science*, in “Public Understanding of Science”, 1 (3), pp. 281-304.
- Yearly, S. (2000) *Making systematic sense of public discontents with expert knowledge: two analytical approaches and a case study*, in “Public Understanding of Science”, 9 (2), pp. 105-122.

* * *

Open Data in the Private Interest

Meg Young

Twin goals of the open data movement

Open data embraces a vision of public participation, collaboration, and transparency. At the same time, it is intended to foster efficiency in government via private enterprise competition and innovation. To the extent that these twin goals are expressions of those behind the open data movement, they parallel a tension rooted in the liberal tradition between freedom and property (Coleman 2013); a tension which manifests at the fault lines of the Free vs. Open Source (F/OSS) software development communities.

Whereas Free software licenses collaborative work with no restrictions on distribution, modification, or use, Open Source development products are proprietary. Nathaniel Tkacz traces the roots of the ‘open government’ movement to an open source rationale, as it is “business-backed” and “compatible with a new form of capitalist accumulation” (Tkacz 2012, 393, 395). While the open data movement is conceptually distinct from its open government and open source kin (Schrock 2016), its conceptual heritage carries the same tension between egalitarian collaboration and private sector innovation. Here, I draw from observations in my fieldwork in Seattle, one of the earliest municipal open data programs in the US, to surface how the private sector has shaped its design and execution.

Open data programs are consistent with longstanding neoliberal goals to make government more efficient by applying market logic to government and embracing disaggregation, in an effort known as New Public Management (Longo 2011; Bates 2012). From the outset, President Barack Obama claimed that open data can foster cross-sector collaboration “with nonprofit organizations, businesses, and individuals in the private sector” (The White House 2009). Seattle’s CTO cited two goals for its new open data program, saying, “With new, constrained government budgets, we’re able to leverage a large community of people outside government to make government... more accessible to everyone” (Socrata 2010). In both cases, open data initiatives were advanced to cut budgets while fostering private enterprise.

At the same time, open data discourse focuses on public engagement and access. To the extent that the private sector is acknowledged in civic hacking discourse, the emphasis is on local software entrepreneurship (Barns 2016) and local social good outcomes (O’Reilly 2013). The ‘civic tech’ space frames the users of open data as local civic hackers and social entrepreneurs (Goldsmith and Crawford 2014; Goldstein and Dyson 2013). From the early days of U.S. open data, a non-profit called Code for America worked with volunteers to promote and build usable, intuitive interfaces on public-facing government services, earning comparisons to the Peace Corps (Wadhwa 2011). Seattle’s local brigade convenes data enthusiasts to work on projects with a local focus (Young and Yan 2016). In turn, Seattle’s municipal government embraced a public engagement strategy, creating a position for a Civic Technology Advocate to encourage local citizen participation. Even as municipal open government data initiatives promulgate a discourse of data in the public interest, the ‘public interest’ is defined in a way that circumscribes private companies and the monetization of public data. Indeed, a primary goal of Seattle’s open data program was to stimulate the local economy in the wake of the global financial crisis (interview, Jan. 23, 2017).

Uses and users of open data

Neither civic hackers nor local entrepreneurs are the primary users of

open data—large companies are. Companies like Zillow and Yelp commercialize civic data, which trade literature refers to as “unlocking its value” (Manyika et al. 2013). The data broker industry, including companies such as Acxiom and Experian, are under-represented in open data discourse relative to their outside use of it (Federal Trade Commission 2014). Recall that Marquet (this issue), describes volunteers who make open and collaborative map data “bewildered” one day to discover Google Maps had been hired by another SNCF subsidiary to do the same work. This moment illuminates vying interests in the production of municipal data, and local civic hackers’ surprise to find themselves competing with such behemoths.

To be clear, private interests in open data do not preclude social good outcomes. Many private sector partners perform an important role to make government data more accessible, usable, and valuable to municipal residents. For example, New York City based company SiteCompli tracks local regulations, inspections, and violations, helping developers keep their properties safe and up to date. Other companies use open data to unjust ends, such as

charging for access to otherwise open data, or using open data to compile dossiers on individual residents. The data broker industry generates billions of dollars a year (Federal Trade Commission 2014).

Corporate open data users describe themselves as intermediaries, working on behalf of the public to derive value from otherwise inscrutable raw data assets. In his comments to public sector personnel at a ‘customer summit,’ one CEO commented, “You need our participation to effectuate the changes you are trying to make” (Renninger 2015; Socrata Customer Summit video 2014). The speaker goes on to provide the following diagram of the ‘open data triangle,’ in which private facilitators (and Socrata) “tak[e] data [from government] transform it, and provid[e] real value” (Ibid.) Many Seattle employees share this perspective; arguing that sharing data allows municipal governments to “better to focus on our strengths and let Google figure out how to get people around town” (interview, March 10, 2015). Here, I take a closer look at Seattle’s open data platform host, Socrata, to surface differences between civic hackers and the private sector as open data users. Socrata is has a private-sector, proprietary software-as-a-service solution for hosting government data. It hosts hundreds of open government data programs, and provides a suite of web tools for user-friendly data analysis. The City of Seattle pays Socrata an annual fee to run its Open Data Platform, data.seattle.gov (known locally as “DSG”), with optional add-ons for data visualization services like the Open Budget application (Levine 2017).

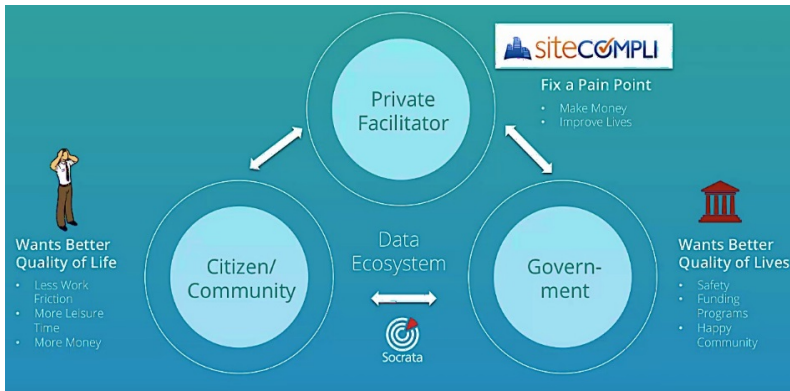


Figure 3 – Socrata Customer Summit 2014 video; Slide from presentation by CEO of SiteCompli, Ross Goldberg

Vendor services adopted for pragmatic reasons have unintended consequences in that they are private sector entities serving a public role. At the time of this writing, there is only one person from Socrata’s 200-employees whose job is to answer requests from civic hacker users, even as the company serves more than 100 municipalities. A focus group with local civic hackers describes Socrata as a barrier between Seattle residents and their government:

A market niche has appeared of intermediary companies... These guys are now our front-end, and they merely shifted it to a closed [one], and it being a closed corporate model actually exacerbates [access issues], because then there's little ability to influence the scheduling of those projects or even the technical capability. So, I have a fear that those intermediaries will inadvertently become a larger barrier than dealing with a government agency that I can always hit with a Freedom of Information Act [request]. You know, pound on the desk—I'm a citizen!" (Focus group, Code for Seattle February 12, 2015).

Rather than understanding industry as a facilitator to public uses of open data, this respondent understood it instead to be an additional interloper, if not a barrier.

Civic hackers also feel limited by the suite of tools and functionality on Socrata’s platform, and find data quality issues (Young and Yan 2017). In this volume, Hoyng (this issue) anticipates these challenges, saying that:

Although the Open Data discourse hails transparency as a democratic-political value, the protocol is not positioned as a right but rather as a service, meaning it exists at the state’s discretion.

Under resource constraints, open data platforms must make choices that have consequences for what their intended uses and users will be. While Socrata has made efforts since 2015 to improve the usability of the DSG platform, it has advanced an initiative in the meantime to open a parallel solution, targeted toward meeting the needs of commercial users.

Private company users of the open data network

This latter platform is called the ‘Open Data Network’ or ODN. Open Data Network is a strategic partnership between Socrata, Yelp, Zillow, SiteCompli, and other companies to make open government data more amenable to enterprise uses. It does not cost any money to partners or municipal governments for their data to be used in this effort. Since 2015, Open Data Network has morphed into something akin to a public-facing search engine for open data, which will eventually index all open data available across jurisdictions, especially that of Socrata’s customer governments. A state employee who leads its open data program sees ODN as a “huge” value add for public agencies, in that it increases the findability (and usage) of their open data (interview February 15, 2017). The long-term vision for the project has evolved from a ‘search engine’ to an interface akin to WolframAlpha that can answer natural language queries. Most crucially, the partnership generates cleaned and standardized data assets to make them easier for enterprises to use across jurisdictions.

ODN centralizes data for large companies to pull this standardized data from a single source. In making it interoperable across jurisdictions, Socrata situates itself as the obligatory passage point in its partners’ enterprise data use (Callon 1984, Söderström et al. 2014). Courmont (this issue) finds that actors that consolidate data foster a new locus of power; “the consolidated data is a boundary object allowing the coordination of various actors through a common representation of the urban space. The consolidation gives the producer new regulation opportunities by gathering these actors around his data” (Courmont this issue). An employee at Socrata explained that such standardization and reach will facilitate ODN partner companies to expand across geographic markets. However, the transformations and cleaning done to the data to prepare it for ODN are not synched back to customers’ own platforms, like DSG, the primary means by which Seattle’s civic hackers access data.

Through a platform studies lens, Socrata’s division of its services into two open data platforms indicates this divergence in the uses and users it serves. Van Dijck (2013) combines political economic and Actor-Network approaches to examine how the design of platforms influences users and content. Adapting van Dijck’s approach helps us to view Socrata with a critical eye towards the distinct missions of data.seattle.gov and the Open Data Network. Her work surfaces urgent questions about platforms’ ownership, governance, and business models.

Relationships between public agencies and partnerships like ODN have

persuasive power. Recall another case study in this issue, the Open Street Maps (OSM) partnership with Transilien, which is described as a “Trojan Horse” - “a nice way to easily open up the doors of municipalities” (Marquet this issue). Similarly, ODN has fostered an initiative to set open standards for local governments. These standards specify the structure, metadata and formats in which housing sector data would be more useful to partners like Zillow (Renninger 2015). Few competitors are participating in this effort; Zillow thus gets an amplified voice in the types of data that cities should be releasing, and the standards that will apply. In the UK, Bates (2012) similarly found that open data was produced at “marginal cost (generally zero)” to provide a marketable asset to private industry. Privately produced standards for open data have rhetorical pull with public agencies, which are purposed to “unlock” local economic value for companies and residents via the release of machine-readable data.

Insofar as the Open Data Network is a nascent municipal open data standards organization, it is advantageous to participating companies. Busch points out how the process of standards-making is also a type of power:

However much standards appear to be neutral, benign, merely technical, obscure, and removed from daily life, they are, I argue, largely and unrecognized but extremely important and growing source of social, political, and economic relations of power. Indeed, in our modern world, standards are arguably the most important manifestation of power relations... [which are] present only when [they are] performed or enacted (Busch 2011, 28).

Standards indeed increase the usability and interoperability of multiple jurisdictions’ data, but they also shift the labor of making data usable from within Zillow – which previously had to clean or standardize data it takes in – to the workforce within municipal governments. This case provides evidence of “the difficulty for keeping standards for things and those for people apart;” changing standards data publication re-configures personnel, labor, and organizations in turn (Busch 2011, 26).

The power of ODN lies in its ability to set priorities and informally lobby its customers to spend resources on opening data that partners find valuable, such as real estate data, via the domains it chooses to release standards. Municipal data is not frictionless to open (Denis and Goëta 2017); it must be collected in a machine-readable format, assessed for risk to privacy and liability, redacted where needed, curated via metadata and data dictionaries, and sent to Socrata’s intake system (and updated manually or automatically). This labor and time may be directed into any number of open datasets that would be useful for research, social justice, or improving public services. However, ODN could persuade governments to emphasize commercializable datasets for release at the expense of others. As governments move forward with open data programs, a greater appreciation of the distinct public and private interests in open data will help to

make more purposeful decisions about which datasets to open, and to what end. Given the labor, resources, and time that governments dedicate toward preparing datasets for publication (Denis and Goëta 2017; Courmont this issue), these resources should be expended with a clear idea of the intended outcomes in mind.

Acknowledgements

I would like to thank the editors of this issue for their valuable feedback. I also thank Antoine Courmont for his insights, especially that the ODN is an obligatory passage point. All errors and omissions are my own.

References

- Barns, S. (2016) *Mine Your Data: Open Data, Digital Strategies and Entrepreneurial Governance by Code*, in “Urban Geography”, 37 (4), pp. 1-18.
- 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*, in “The Journal of Community Informatics”, 8 (2).
- Busch, L. (2011) *Standards: Recipes for Reality*, Infrastructures Series, Cambridge, MIT Press.
- Bogost, I. and Montfort, N. (2009) *Platform studies: Frequently questioned answers*, in “The Proceedings of the Digital Arts and Culture Conference”, December 12-15, Irvine, CA.
- Callon, M. (1984) *Some elements of a sociology of translation: domestication of the scallops and the fishermen of St. Brieuc Bay*, in “The Sociological Review”, 31 (1), pp. 196-233.
- Coleman, G. (2013) *Coding Freedom*, Princeton, Princeton University Press.
- Denis, J. and Goëta, S. (2017) *Rawification and the careful generation of open government data*, in “Social Studies of Science”, pp. 47 (5), pp. 604-629
- Federal Trade Commission (2014) *Data brokers: A call for transparency and accountability*, Washington, DC.
- Goldsmith, S. and Crawford, S. (2014) *The responsive city: Engaging communities through data-smart governance*, Hoboken, John Wiley & Sons.
- Goldstein, B. and Dyson, L. (2013) *Beyond transparency: Open data and the future of government*. San Francisco, Code for America Press.
- Levine, T. (2017) *Semi-open data about pricing of open data*, <https://thomaslevine.com/!/open-data/socrata-products/> (retrieved July 14, 2017).
- Longo, J. (2011) *#OpenData: Digital-Era Governance Thoroughbred or New Public Management Trojan Horse?*, in “Public Policy & Governance Review”, 2 (2), pp. 38-51.
- Manyika, J., M. Chui, P. Groves, D. Farrell, S. Van Kuiken, and E. A. Doshi. (2013)

- Open Data: Unlocking Innovation and Performance with Liquid Information*, McKinsey Center for Government and McKinsey Business Technology Office.
- O'Reilly, T. (2013) *Open Data and Algorithmic Regulation*, in B. Goldstein and L. Dyson (eds.), *Beyond Transparency*, San Francisco, Code for America Press, pp. 289-300.
- Renninger, M. (2015) *The New Internet: Open Data Network*, <https://socrata.com/blog/new-internet-open-data-network/>. (retrieved July 14, 2017).
- Schrock, A. (2016) *Civic hacking as data activism and advocacy: A history from publicity to open government data*, in "New media & society", 18 (4), pp. 581-599.
- Socrata Customer Summit 2014 (2014), https://www.youtube.com/playlist?list=PLI8oOhEXXuPscwXyJbdVoMtvswqfc9_tZ (retrieved July 14, 2017).
- Socrata (2010) *How Seattle Built an Economy of Service*, <http://web.archive.org/web/20170113055135/https://socrata.com/stories/city-seattle/> (retrieved July 14, 2017).
- Soderstrom, O., Paasche T. and Klauser F. (2014) *Smart cities as corporate storytelling*, in "City" 18 (3), pp. 307-320.
- US Federal Trade Commission (2014) *Data Brokers: A Call for Transparency and Accountability*, Washington, DC: US Federal Trade Commission.
- Van Dijck, J. (2013) *The culture of connectivity: A critical history of social media*, Oxford, Oxford University Press.
- Wadhwa, V. (2011) *Code for America: An Elegant Solution for Government IT Problems*, in "The Washington Post", Dec. 19.
- The White House (2009) *Memorandum for the heads of executive departments and agencies, Transparency and Open Government*.
- Tkacz, N. (2012) *From open source to open government: A critique of open politics*, in "Ephemera", 12(4), pp. 386-405.
- Young, M. and Yan, A. (2016) *Civic Hackers' User Experiences and Expectations of Seattle's Open Municipal Data Program*, in "Proceedings of the 50th Hawaii International Conference on System Sciences", pp. 2681-2690.

