Where's the Database in Digital Ethnography? Exploring Database Ethnography for Open Data Research

Ryan Burns
University of Calgary

Grace Wark
Alberta Wilderness Association

Abstract

Contemporary cities are witnessing momentous shifts in how institutions and individuals produce and circulate data. Despite recent trends claiming that anyone can create and use data, cities remain marked by persistently uneven access and usage of digital technologies. This is the case as well within the emergent phenomenon of the "smart city", where data increasingly constitute a fundamental dimension of urban life, governance, economic activity, and epistemology. The digital ethnography has extended traditional ethnographic research practices into such digital realms, yet its applicability within open data and smart cities is unclear. The method has tended to overlook the important roles of particular digital artifacts such as the database in structuring and producing knowledge.

In this paper, we develop the database ethnography as a rich methodological resource for open data research. This approach centers the database as a key site for the production and materialization of social meaning. The database ethnography draws attention to the ways digital choices and practices - around database design, schema, data models, and so on - leave traces through time. From these traces we may infer lessons about how phenomena come to be encoded as data and acted upon in urban contexts. More specifically, we argue that open databases limit data types, categorize and classify data to align with technical specifications, reflect the database designer's episteme, and (re)produce conceptions of the world. We substantiate these claims through a database ethnography of the open data portal for the city of Calgary, in Western Canada.

Introduction

Contemporary cities are witnessing momentous shifts in how institutions and individuals produce and circulate data, raising pressing new questions about the ways in which data, social relations, and urban environments are mutually constituted. Despite recent trends claiming that anyone can create and use data, such as in conversations around Big Data, social media, and user-generated data, cities remain marked by persistently uneven access and usage of digital technologies. This is the case as well within the emergent phenomenon of the "smart city", where data increasingly constitute a fundamental dimension of urban life, governance, economic activity, and epistemology (Gaffney and Robertson, 2016; Luque-Ayala and Marvin, 2015; Townsend, 2013).

Within the emergent research agenda on smart cities, researchers have begun to connect the roles of open data platforms – databases provided by formal government agencies in the name of

"transparency", "accountability", and "innovation" – to claims of a city's "smartness" (Barns, 2016). Open data platforms serve as a key tactic by which the smart city is *known*; they mobilize a dominantly realist epistemology that obscures the politics and contestations around how people, places, and ideas come to be encoded as data (Author, Date; Kitchin, 2014).

At the same time, research within science and technology studies (STS) has long accepted that technologies embody social, political, and economic values and imperatives (Latour, 2000; Wajcman, 1991; Winner, 1985). By extension, examining a technology helps to cull insights into social processes. In recent years this principle has been applied to *data* and its supporting technologies, such as databases, sensors, software, and categorization schema (boyd and Crawford, 2012; Author, Date; Kitchin, 2013). As Dalton and Thatcher (2014: n.p.) note, a data model "structures and encodes information in one way or another according to the visions of the team of data engineers, scientists, and developers that created it." However, while considerable research has elucidated such social and political dimensions of smart cities *per se*, very little empirical work has looked into the ways open data platforms shape knowledge of the city and its inhabitants.

Even less research has interrogated the affordances and limitations of particular methodologies for understanding open data platforms as strategies of smart cities[1]. Gray (2014) has made perhaps the most prominent contributions in this regard, advocating for an approach toward open data that utilizes the Foucauldian genealogy. However, it is most often the case that researchers bypass discussing their methodological approach entirely (Barns, 2016; i.e., Sieber and Johnson, 2015). In cases where the researcher makes their methodology transparent, they have typically translated "offline" methods to the new arena, such as in case studies leveraging interviews, and content or discourse analysis of policy documents. This is despite acknowledgement that "research would … benefit from … in-depth interviews with all actors, tracing the interconnections and interdependencies of the various actants, *and deconstructing the attendant code*" (Kitchin et al., 2015, emphasis ours) of open data platforms and dashboards.

This raises the pressing question facing researchers of smart cities and open data: how does the new socio-technical assemblage of smart cities and open data call for new – or at least adapted – digital methods? If we are to take the digital as an active space co-constituted by social and urban relations, as we contend below, rather than as a simple context, then current methodologies overlook a critical actant in the production of urban knowledge. We address this question partly by drawing into critical scrutiny such taken-for-granted aphorisms as "[d]ata do not speak for themselves" (Dourish and Cruz 2018, 1); while we do not claim that the "language" of data is straightforward or that all who listen will hear the same thing, we think that data *do*, in a sense, *speak* – and that by ignoring that "speech", scholars miss a key opportunity for culling insights into smart cities and open data. In other words, extant methods transposed from "analogue" research approaches risk obfuscating the very active role that the digital plays in constituting urban knowledge, and this risk is accentuated in the smart cities context, given that in these contexts digital technological infrastructures are deployed precisely to impact human behavior. It is to this point that we contribute in this article.

---

[1] By this, we do not mean the sorts of methodologies – and by extension epistemologies – smart cities and open data platforms proffer, but rather the ways in which researchers can systematically and rigorously understand the production and implications of open data and smart cities. In this way, Kitchin et al's (2015) poignant argument that open data and data dashboards represent a realist epistemology comprised of fact-finding methods, and Mattern's (2013) provocative critique of what she calls "methodolatry" are important, but irrelevant for the questions at hand.

Below, we argue that the new digital milieu of open data platforms within smart cities requires new methodological resources, and we show that the *database ethnography* contributes rich understanding to these phenomena. The more common methodology of digital ethnography moves in an informative and important direction. However, in contrast with digital ethnography, database ethnography recognizes that data infrastructures *themselves* shape human behavior and conceptualizations of the world. By extension, scholars may glean insights into social process by observing the ways such infrastructures frame knowledges and behavior, and the ways they translate phenomena into data – applying these principles to a smart cities context draws these contributions in stark relief. In what follows, we first describe the database ethnography method, drawing on and still distinguishing this method from digital ethnography. We then describe the empirical case to which we applied the database ethnography in order to understand its affordances and limitations. After reflecting on the outcomes and particularly the benefits of the method application, we offer a set of nuances and caveats related to the database ethnography.

2. Digital and database ethnographies

The emergence of smart cities and open data platforms brings to the fore methodological debates and attendant epistemological considerations that have motivated qualitative research for decades. Within this time, a growing interest in digital research methodologies has fostered insightful debates around the relevance of ethnographic methods for internet-based research (Murthy, 2008). Many diverse approaches to ethnography are in practice, but most draw on extended embodied experience with a community in their environment (Atkinson, 2015). Ethnography relies on the understanding that "social structures are reproduced and challenged through the everyday processes of social life" and that "intersubjectively constructed sets of meanings [encode] these everyday practices" (Herbert, 2000: 553). For some commentators, the digital sphere writ large signified a disembodiment that ran askew to this presumed directness and physical interaction of ethnography (Gobo, 2008; Hammersley, 2006). This reflected common early conceptions of the Internet merely as a place-less facilitator of disembodied interactions – a problematic conception that obfuscated the inherently social and political nature of digital interactions (Nakamura, 2002). In fact, in their pioneering ethnography of Trinidadian use of the Internet, Miller and Slater (2001) found that online interactions with the Internet and with each other on the Internet were both shaped by "offline" cultural and social norms and values. Still, the digital presents a number of challenges to traditional ethnography. The question for many ethnographers was not whether there is a need for "a distinctive and separate methodological strategy that is 'digital'" (Atkinson, 2016: 236). Rather, the digital was seen as a productive "site" due to the ways in which digital technologies mediate the production of social relations and social meanings.

The rise of digital ethnography (Hine, 2000) made an important theoretical contribution: it posits that social life, micro- and macro-logical structures, and social meaning are all (re)produced, captured, internalized, and asserted in online spaces (Murthy, 2008). In her early work on "virtual ethnography", the precursor to the more common contemporary phraseology used here, Hine (2000) notes that the Internet serves not just as a site for new communities to form, but that it also implicates offline interactions and relations such as class, knowledge legitimacy, and the proliferation of cultural artifacts. Thus, digital ethnography studies social interactions occurring in online spaces, but also how digital technologies shape people's interactions with each other and their environments. Most researchers tend to approach digital ethnography, then, as traditional ethnographic methods adapted to the context of Internet and web-*mediated* communities.

Ethnographers have explored these principles in a wide range of application areas: social media (Hinton and Hjorth, 2013), chat rooms (Mileham, 2007), online commercial spaces (legal and otherwise) (Barratt and Maddox, 2016; Robinson and Schulz, 2009), gaming sites (Boellstorff, 2015), gambling "new media worlds" (Farnsworth and Austrin, 2010), and many more. To date, digital ethnographers have insufficiently dedicated attention to the database itself as a space for the production of social meaning. The database is, of course, a cornerstone digital technology for smart cities and open data, and the utility of extant methodological resources is therefore unclear.

This omission relates to a key consideration confronting digital ethnography: how does the digital ethnographer contrive "the field" in which they embed themselves? This consideration has implications far broader than purely methodological, instead speaking to the way in which the researcher conceives of the nature of sociality (Nast 1994; Walker, 2010). It is an inherently political question involving the privilege and power of the ethnographer to draw such boundaries and to consider themselves "in" the field and yet not "of" it – a duality that is at once problematic and inherent (Tunçalp and L. Lê, 2014). All researchers produce a "field" of study that they purportedly are free to exit at their desire (Massey, 2003). For digital ethnography, the drawing of this boundary around "the field" is no less problematic, but is no more problematic than in traditional ethnography, either (Beaulieu, 2004). Whereas Forte (2004) argues that the field – or the "site" – "pre-exists the ethnographer and … the ethnographer comes to 'visit' as an 'outsider'," we contend the opposite: that the site is produced in the act of the ethnographer defining it. This is not to negate communities' claims to coherence and collective action, but rather to draw attention to the performative and contestable nature of claiming a "field". This reflexive uneasiness with delimiting "the field" informs recent nascent conversations about *database ethnography*, wherein the field is construed as the database itself.

Database ethnography can be seen as a branch of digital ethnography that takes as its object of study – as its *site* – the under-examined technological frameworks that influence how phenomena are captured and represented as data (Schuurman, 2008). Database ethnography calls into question the assumptions, categorizations, social lineage, epistemological frames, and consequent implications of data models and their storage, representation, and dissemination. To a degree, it is interested in the linkages between database changes and parallel changes in perceptions and ways of knowing, drawing on Geiger and Ribes' (2011) insistence that ethnographers attend to the "traces" left by past versions of the database.

In her elaboration of the method, Schuurman (2008) is primarily interested in understanding the nuances, vagaries, and languages of data production, but only as a means to the end of improving metadata. If scholars and researchers can "uncover" the conditions and contexts in which data were produced, this will enable them to write richer explanations of why and how data can be used. For example, different provinces or states may have differing criteria for linking suicides with gambling-related activities, leading to unexpectedly different numbers across legal jurisdictions (Schuurman, 2008).

Contra Schuurmann, we argue that database ethnography can be useful for elucidating social process as its primary aim. In other words, the value of a database ethnography need not be the way it extends and improves metadata. Database ethnography, like digital ethnography and "traditional" ethnography, can be engaged to illuminate the ways in which meaning is produced, the ways individuals and groups internalize and resist larger social structures, and the ways humans seek to

interpret their environments (Beneito-Montagut, 2011; Thomer and Twidale, 2014). We concur with Hine (2006) that the database does not produce deterministic outcomes or wholesale transformations of practices; at the same time, borrowing from the social construction of technology debates (Pinch and Bijker, 1987; Star, 1999; Wajcman, 1991; e.g., Winner, 1985), the database reflects values, norms, epistemologies, social relations, and power that in turn influence how people interact with the world and each other. Similar processes have been noted in classification systems – both digital and analogue (Bowker and Star, 2000; Author, Date). Indeed, in an ethnography of scientific collaboration, Bietz and Lee (2009) found that the database served as a point of rearticulating knowledge. Rather than pre-determining conceptual frames, or speaking differently to different specialty groups, Bietz and Lee argue that databases serve as a locus for asserting knowledge frames – the proper ways of knowing a particular phenomenon – and for rearticulating those frames in contestable ways.

These discussions underscore that knowledge is often characteristically contextual, contested, and tacit (Collins, 2001). By extension, the database, which seeks to encapsulate people, places, phenomena, and knowledge as data, captures historically and geographically-specific ways of understanding the world. Further, it captures the ways individuals and groups make sense of their environments. Borrowing from Manovitch (1999) and Simondon (2017), the ethnographer can take the database as its site simultaneous to its conception of the database as a social and cultural artifact. In the nascent context of open data research within smart cities, greater attention to the *database* is needed than what digital ethnography has tended to give it.

The research on which we report here questioned the ways open data databases shape human behavior and conceptualizations of the world. Like most databases, open data are comprised of data models, classification schemes, multiple file format download options, and structured API access. They serve as public-facing repositories of official government datasets. Open data are thus best understood as information infrastructures that are situated within broader information ecosystems and socio-technical relationalities (Bowker, 2000; Star, 1999). Methodologically, then, it is important to understand the ways social, cultural, and political norms and values are encoded in persistent technological artifacts, within larger (nominally "offline") ways of understanding the urban world. To do so requires that we look not only at individuals' digital presence and communities interactions online, but instead turn toward the influence the database *itself* has on the production of these modes of being and knowing. We thus argue that new methodological resources are necessary to understand this new mode of sociality, and that the database ethnography considerably advances this goal. It is to this project that we now turn.

3. Database ethnography on Open Calgary

The objective of this research was to explore the unevenness, epistemological claims, and political purposes of open data platforms as a strategy of smart cities. For the remainder of this article, we show the utility of the database ethnography for achieving these goals. We approach this goal with the assumption that a technical object, in this case the database, constitutes an artifact that solidifies and materializes social, cultural, and political relations and processes. The database captures and internalizes these relations and processes in its design, its coverage, its data models, categorization schemes, table structures, and more. In other words, the database never exists outside social and political relations and processes. After it is implemented, it then impacts human behavior, knowledge, and sociality. By extension, researchers may learn about these relations and processes by

interrogating the database itself. To do so, however, necessitates looking not only at the database itself, but also the decision-making practices, the deliberations, the contexts, and the political-economies within which the database was constructed and within which it operates. To paraphrase Smith (2014, 198), interrogating the database involves observing the systems of meaning that are inscribed in technical artifacts in order to bring data into existence. We thus combined our exploration of Calgary's open data platform Open Calgary with complementary ethnographic methods. In this, we asked pertinent questions around the ways in which researchers can glean insights into social and political relations and processes by interrogating an open data platform. In other words, this research gauged the viability of database ethnography for exploring open data platforms.

Mapping Open Calgary, Diagramming Society

In 2012 the City of Calgary launched its open data catalogue, which was comprised of a small list of machine-readable, downloadable datasets that the city had released to the public. The datasets largely pertained to municipal roles and services, such as neighborhood boundaries, hydrology, census data, and infrastructure. They were made available for download in multiple formats such as in tabulated, geographic, or hyperlink form.

This effort came on the heels of growing interest in Western cities toward open governance and transparency, and data-focused civic engagement. The federal government in Canada had released its open data site earlier that year, "to make Canada one of the best places to do business by ensuring that Canadian entrepreneurs, researchers, academics and voluntary organizations have access to government data in useful formats to help foster innovation, job creation and community services for Canadians," according to then-President of Canada's Treasury Board Stockwell Day (Canada NewWire, 2011). Similarly, in the United States, the Obama administration launched the federal open data platform Data.gov in 2009 (Shueh, 2016) and in a 2013 executive order, mandated that with some exceptions federal data should be made publicly accessible "by default" (Obama, 2013). The executive order focused primarily on benefits for the private sector, such as in the telling justification, "Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process." From the outset, then, moves toward "transparency" were simultaneously motivated by interest in entrenching particular discursive connections between data and economic value.

In 2016 the city replaced its open data catalogue with an open data *platform* managed through the Socrata suite of databases, software code base, application programming interface (API), and user-directed interface. Socrata, a Seattle-based private business, is a dominant platform in the open data field, with clients – primarily governments – spanning globally (Levy, 2016). For Calgary, the shift to an open data platform changed the user interface, but also expanded the website's capacities to include functionality like interactive visualization, API access (rather than direct download), dataset tagging (short descriptors of the datasets), and discussion boards. This was distinct from, yet complemented, two contemporaneous developments: the city's 2015 creation of a "citizen dashboard" that delivered city staff-produced visualizations around self-reporting measures, for example; and the 2014 execution of a new citywide digital strategy implementing new "…trends in digital and open government" (Sevigny and Angelo, 2014). Thus, Open Calgary was released within a milieu of dramatic shifts in the ways city staff envisioned the role of technology and data in their governance practices.

Data collection

Our database ethnography consisted of digital methods supplemented with well-established ethnographic approaches. The ethnography entailed establishing long-term, and ongoing, institutional relationships in which we have embedded ourselves, and which open avenues for more rigorous excavation of Open Calgary. Our ethnographic practice spans across five broad sets of communities: the city administration, local non-profit organizations, community associations, the Open Calgary platform, and our own research team. As reflexive researchers, we are particularly attuned to the ways in which our presence as researchers "in the field", as well as our positionality and interests, percolates influence throughout every aspect of the research project (Berger, 2015; Burawoy, 1998; England, 1994). We focused our attention on those who currently engage or have recently engaged digital data in their work, and developed these connections through partnerships between our research team and our various communities. Embedding ourselves into the digital operations of these groups further entailed in-depth interviews, participant-observation in community association meetings and city council sessions, and ethnographic field note-taking, all of which are described in more detail below. Combined, these methods constituted our primary methodology of database ethnography of Open Calgary.

Database ethnography recognizes that data captured in a database are abstract in several senses of the term. As Schuurman (2008) notes, they are stripped of the social and institutional needs driving data collection, removed from the situations in which someone decided to collect those data, and from the choices about how to represent the phenomena through data models. Database designs, data models, categorization schemes, attributes captured, and more, all mobilize tacit knowledge that actors do not record explicitly in the database. For these and other reasons, database ethnography supplements database retrieval and archival work – work on the database itself – with systematic knowledge production that occurs *outside* the database. We expand on these ideas in the following sections.

Database archive

One of the key strengths of the database ethnography is its raising *digital traces* to the analytical fore. We tackled this goal with an archiving method as a sort of versioning, or snapshots of the platform's collection of data at a point in time. We created a Python script that uses Open Calgary's API to make semi-regular archives, which we supplement with occasional archives of Open Calgary's Citizen Dashboard front-end. The latter involved simple screenshots to capture the types of projects and impacts that the initiative highlights (for an example of the type of information foregrounded there, see Figure 1). Our goal for the Citizen Dashboard was to monitor the way its administrators promote its offerings as components of a broader move toward citizen engagement; therefore, we were not interested in the actual values for the metrics themselves, but rather the ways the metrics were figured into the site.

Figure 1. Calgary's Citizen Dashboard is where the city disseminates some of their visual analyses of open data offerings

Interviews and participant observation

In this research, we were particularly motivated by the assumption latent in existing database ethnography literature that the database exists at the terminus – consistently re-negotiated – of a plethora of social and political processes. By extension, while the database captures and reflects these processes, one must explore the socio-political processes and practices themselves to understand the database. To get there, we conducted 15 in-depth, semi-structured, in-person interviews approximately an hour in length. These interviews aimed to enrich our understanding of the institutional and contextual forces affecting data practices. In these interviews, we asked about the multiple factors and dimensions of the city's decision-making practices in establishing, populating, updating, and acting upon Open Calgary. We hoped to understand, as well, the ways these decisions impacted their use and the ways actors envision the phenomena captured as data, so we also conducted these interviews with community associations and non-profit organizations in Calgary. All interviewees were in managerial and leadership positions, and not all interviewees had interacted with Open Calgary.

We conducted participant-observation activities in collaboration with the stakeholder groups we have identified above. These entailed participating in community association board meetings, contributing to the Open Calgary discussion board, attending data and innovation-related events such as the local "Data for Good" hackathons and social gatherings, promotional events for city infrastructure projects that had engaged digital technologies for citizen feedback, and local data-

related conferences. In these activities, we recorded daily ethnographic field notes throughout the entire research project. In these notes we reflected on research experiences, initial impressions of interviews and participant-observation, linkages with existing research literature, and on our changing conceptions of the open data ecosystem (Emerson et al., 2011; Hsu, 2017).

Unexpectedly, early in this research we were made aware of a University of Calgary grant supporting the development of a university-wide research platform. With our interest in the gaps of open data, we decided to apply for this grant to develop a data sharing platform that captures and disseminates datasets outside the purview of Open Calgary. A description of this grant application is beyond the purposes of this paper, but importantly, it shaped the ways we reached out to stakeholder groups and influenced the ways in which we couched the additional value of partnering with our research project. Namely, we began reaching out to organizations partly by framing our inquiries around their answers' application in the proposed platform. On the one hand this significantly broadened the appeal of our research to our stakeholders; on the other hand, at times we sensed a utilitarian logic percolating through our discussions with them. This has become even stronger after the project successfully received this grant and we began developing the platform. Rather than negating or combating this influence, however, we see it as exemplary of how researchers conduct qualitative research within a technological milieu, wherein they tack between technical pragmatism and theoretical abstraction.

4. For Database Ethnography

Conceiving of the database as a socio-technical artifact can illuminate the unique ways databases structure social knowledge in the smart city. In this section we outline the insights that emerged as a result of our use of the database ethnography. We detail these insights as research findings, but we envision this section as showcasing what sorts of insights are conceivable with the use of our methodological contributions.

Open Calgary functions as an epistemological tool in many important ways. As a tool facilitating particular epistemological claims within a large institutional clearinghouse for city data, the Open Calgary database is a crucial actant within Calgary's open data and smart city movements. It receives legitimacy through the authority of the city, but also through partnering with local technology-related non-profit organizations that promote Calgary's open data. In Calgary, these include organizations such as Data for Good – Calgary, a local group that hosts "datathons" and "Meetups" to engage the public in Open Calgary. In such events participants produce insights into datasets, using data science approaches; these insights have in the past led to some new datasets being released in Open Calgary. These events also launch civic dialogue around the affordances of specific datasets and analytical approaches. With Data for Good as well as the city-hosted Civic Innovation YYC, a primary goal is to produce discourses about what constitutes "data" and how those data can produce insights into urban-administrative processes. Event conveners often directly access Open Calgary in real-time to demonstrate built-in tools and methods for translating data into what they call actionable information and knowledge. Together, such institutional forces, networks, and technological apparatuses constitute a key site where knowledge is produced, framed, and captured as data within Open Calgary. This makes Open Calgary especially important to interrogate as an artifact of cultural, social, political, and institutional values and norms. Here, we draw out four ways Open Calgary performs its role as a socio-political and epistemological actant: it limits data types, categorizes and classifies data to align with its technical specifications, reflects the database

designer's episteme, and (re)produces conceptions of the world. In substantiating these arguments, we acknowledge that these insights emerged directly from the ways in which we mobilized our version of the database ethnography. This is because the database ethnography brings to the analytical fore the ways in which social meaning and knowledge claims are produced, maintained, enacted, and materialized in the digital object of the *database* per se.

First, Open Calgary delimits the types of phenomena that will be stored as data. For reasons both practical and political, Calgary's database engineers designed it to be able to store limited types of phenomena. Namely, Open Calgary was designed to release current government holdings to the public. Datasets come from different departments within the city government, and after a process of needs assessment, vetting, and licensing, open data staff make the data available on the platform. While the data cover a breadth of topical areas, such as environmental data and demographic data, other topics of political import cannot be included, such as grassroots and citizen-generated data, data of a personal or emotional nature, and data with irreconcilable licensing agreements. These limitations are usually the result of an interplay between political and technological imperatives: regarding citizen-generated data, Graham, a leader in the Open Calgary project told us that although there is "nothing [technical] that would stop a crowdsourced dataset from appearing there," the city is ill-prepared to handle these data sources. This partly explains why, as another staff member Danny revealed, the "Dataset Suggestions" discussion board generates new dataset inclusion ideas yet has not directly led to any additional datasets being released in Open Calgary.

Thus, the database designers' decisions around inclusion – *what should this database store?* – are equally questions about what should be excluded (Author, Date). These exclusionary practices are couched within societal debates about legitimacy, representation, and governmentality, or "knowledge politics" more broadly (Brandusescu and Sieber, 2017; Elwood and Leszczynski, 2012). Open Calgary's data offerings are exclusive to current government holdings. These decisions around inclusion and exclusion constitute a power relation between the government and citizenry, wherein the city frames conceptions of "openness" and "accountability" through the technical apparatus of the open data database. In this way, smart cities' efforts to engage open data do not exclusively promote inclusive civic engagement, and instead shift inequalities and exclusionary politics to this new site of the open data database.

Second, Open Calgary contextualizes the recorded phenomena within its extant data models, table structures, schema, and other internal specifications. As of this writing, Open Calgary classifies all datasets into 11 relatively stable categories. Among the more populated categories are "Government" (162 datasets), "Transportation" (70 datasets), and "Environment" (31 datasets). Datasets are also assigned a "type", which refers to the form of the data – including spatial data, charts, calendars, and documents. The breadth of these types has increased since switching to Socrata; previous types primarily differed by encoding (e.g., SHP, KML, Google Maps, Bing Maps). Further, data are assigned keywords called "tags" in order to draw further lines of similarity across the datasets.

The digital practices of "fitting" datasets into these specifications draws on a combination of the database managers' formal expertise, the default technical specifications of the Socrata platform, and the managers' looking to other cities to ascertain typical approaches. According to Alan, a manager of Open Calgary,

> *"Socrata creates the categories, and then we look at our existing inventory and then begin to identify and determine … this dataset definitely fits under category A. Category B – this dataset definitely fits under that category. … We've kind of followed what other government bodies have done, in categorizing their datasets under those respective themes. … we've just kind of followed suit. ... I mean, we're just trying to keep a fairly consistent theme with what everyone else is doing."*

Additionally, Open Calgary staff have diverse backgrounds, often of a scientific nature, which shapes their placement of data within the database. Despite these complexities, however, most data have been contextualized within the default Socrata classification, types, and tagging schemes. This process, of course, fits phenomena within the epistemological frame of database structures.

Third, and related to the previous point, Open Calgary influences the conception of various phenomena in the world. By this we mean that users' understanding of what constitutes "environmental", "governmental", and "health" data is contested, reified, and altered as users access and interact with data categories and tags. Open Calgary serves as a site for the production of such concepts, categories, as well as the apprehension of phenomena. In this argument we draw upon a vast body of work with disparate theoretical commitments, but all of which presuppose that technical objects fix meaning across contexts (Harley, 1989; Latour, 1986; Simondon, 2017; Star and Griesemer, 1989). Open Calgary can be seen as a site in which the user or audience encounters data categorized in such a way as to make claims about what constitutes "the environment" and other categories. In other words, the "Environment" category (as well as the other categories, types, and tags) does not *reflect* the environment so much as actively construct the concept. This is particularly important as categories shift, especially in the transition to a Socrata interface, where the categories have been decided beforehand by a geographically distant entity. By extension, the smart city and its broad range of actors concretize, within the database, the epistemological frames through which the public explores and knows their city, and through which city administrators evaluate and promote an open database's offerings.

Lastly, building on this, Open Calgary captures the database designer's abstraction and apprehension of the phenomena that the database stores as data. Traces of the database designer are discernable in the categorization scheme, data models, database functionalities and table relations, backend software, interface layout, tags, and metadata. For example, Open Calgary hosts a "Community Boundaries" spatial file that stores the geometric bounds of each "community" along with its name, geographic sector, development status, zoning designation, and numeric identifying code. However, from the data stored in Open Calgary it is unclear the purposes for which the dataset was created, how *community* was defined for the dataset, and the file's spatial resolution, among other qualities. One could read into the dataset that it reflects the creator's best attempts to provide what they believe to be necessary information regarding their conception of *community*. Everything from the spatial resolution at which boundaries were drawn, to the origin for *communities* reflects the creator's claims about what is important and necessary about "Community Boundaries" in the city. By extension, we may assume that Open Calgary staff released the dataset presuming its faithfulness to these qualities make it a useful and important dataset for public consumption; put otherwise, they concur on some level with the creator's abstraction of *communities* within Calgary. In conversation with Open Calgary staff[2], we were pointed to documentation that conflates two meanings of

---

[2] See https://data.calgary.ca/Base-Maps/Community-Boundaries/ab7m-fwn6 for this public discussion.

"community": as administrative units – "[c]ommunity associations come in all shapes and sizes and … [address] local issues, opportunities and needs" – and as interpersonal relations that shape the urban experience – "[c]ommunity associations … create that small town feeling in a large urban centre"[3]. This ambiguity has concrete implications for database design, but more importantly, implies that one may detect the database designer's conception of "community" by looking at the ways it was captured as data in this dataset.

5. Nuances and caveats: performative and reflexive research
Performativity of databases

We approached our database ethnography with an eye toward the database's performativity rather than its ontological status. This involved privileging in our analysis the social and political contexts in which databases construct meaning. In so doing, we draw on Drucker's (2013) insistence that digital objects should be understood as events – with questions around what they *do* – rather than as static objects – with questions around what they *are*. To apply these principles in database ethnography entails a continual process of reflexivity, analytically tacking between the implications of the database as a social intermediary on the one hand, and our own reconfiguration of evidence and data as researchers on the other hand. As we seek to understand how Open Calgary implicates knowledge production and political relations, we repeatedly asked how our presence as researchers shaped the knowledge we produced about the database[4]. Performativity, for the database ethnography, must be conceptualized as a range of practices that include everything from the database itself shaping social life, to the ethnographers' interactions with our subjects, to the process of writing (Hsu, 2017).

Two examples of our performative database ethnography lucidly illustrate the epistemic complications of – and simultaneous need for – a focus on the database within digital ethnography. First, we observed very early in the study that in our participant-observation and our interactions with interviewees, they adapted to our presence as researchers and to the implications of our questions. On one occasion, as mentioned above, we asked our interviewee Graham about the process one would need to take to have a grassroots-generated dataset included in the official database. Graham responded that the city does not have a process in place for such data; later after our interview, however, Graham emailed us a blog entry he had recently read, in which the author advocated for more receptivity to adopting citizen-generated datasets. Graham explained in this email that he has been thinking more about this issue lately, and that our question may have prompted some momentum in that direction. As we progressed through this project, we developed a rapport with city staff members such that they often sought our feedback on Open Calgary and its data holdings. Any questions we asked, our reactions to their responses, and the very existence of our project likely altered their perspective on the open data portal and its successes.

Second, in our research we devoted considerable attention to the role we as ethnographers played in the production of knowledge about Open Calgary, an influence that percolated through all aspects of the project: from forming research questions, to organizing and arranging evidence, to the writing of results (Berger, 2015). Open Calgary mediates knowledge and sociality, but our performances as

---

[3] From https://calgarycommunities.com/faq/#three.

[4] We draw additional inspiration for this notion from Burawoy's (1998) extended case method, another form of reflexive ethnography.

Where's the Database in Digital Ethnography?

ethnographers shape how we come to know that mediation, an idea with well-worn traditions in various strands of STS (e.g., Harding 1987; Latour 2004). This dimension of performativity productively generates what Hsu (2017) calls "partial truths". Just as Open Calgary stores and represents *some* abstractions of the world at the cost of excluding others, our approach toward database ethnography was filtered through our own political and intellectual commitments. Focusing on the database facilitated a reflexivity with implications for the materiality of Open Calgary – we interpreted Open Calgary through our understanding of the uneven, contingent, and often problematic impacts open data and smart cities have across the world. This set of assumptions came into tension with many of our research subjects' optimism about the promises of the database to transform politics and society, effectively shaping the research process, the collected data, and the knowledge we have produced (Nast 1994).

Temporality of digital artifacts

Ethnography has always struggled with questions of temporality. How long should the researcher immerse themselves in a site (Herbert, 2000)? From what resources can we draw to explain social *histories* (Burawoy, 1998)? How should the ethnographer arrange a sequence of events and evidence into a meaningful narrative (Rapport, 2000)? With our database ethnography we encountered unexpected difficulties "tracing" Open Calgary through its past. We relied on staff members' memories, and digital archives of the database, but both were more restrictive than we anticipated. For example, staff do not typically deliver archives to researchers, and many staff members have changed roles or left their positions entirely. Our own archiving constituted a "snapshot" of the database as it was at the time of archival, and so long-term archival work in this direction is likely to expand the principles we can generate through this method. Despite this, many core assumptions of database ethnography proved reliable: Open Calgary betrays hints at its historical structure and functionalities development, it frames and delimits the translation of knowledge and experience into particular forms of data, and it serves important social and political roles. Considerations of temporality within digital ethnography, in principle, translate well into a detailed focus on the database, as institutions often retain past versions, and decision-making practices directly impact the materiality of the database itself (i.e., its categorization schemes, API functionality, etc.).

Still, while we argue that the database ethnography is a useful exploratory and analytical tool for digital platforms that store some sort of data in databases, we acknowledge that not all platforms do so. Some digital platforms store little – if any – data within a formal database structure. We self-consciously draw attention to the database artifact due to its strong alignment with ethnographic approaches, asking, "Where is the database (and other digital artifacts) in digital ethnography?" Yet we simultaneously recognize layers of politics involved in other socio-technical configurations, such as the digital platform (Graham et al., 2017; Srnicek, 2017), the software stack (Bratton, 2015; Author, Date), the algorithm (Crampton and Miller, 2017), and others. In some ways similar to these other digital objects, the database's temporality is salient in its versioning systems, persistence of data models and schema, and technological "closures" (Author, Date). We thus submit that for digital ethnography, the researcher can glean important insights into social processes by looking at the database *per se*.

6. Conclusion

In this article we have argued for the renewal of the database ethnography within digital research, and in particular within open data and smart cities research. The phenomena within these areas of

inquiry take place largely in and through the digital artifact of the *database*, which correspondingly encapsulates social systems of meaning-production. As the database changes through time, it leaves "traces" behind that may contribute insights into how knowledge about the smart city is framed, produced, contested, curated, and represented. To date, researchers have not paid sufficient consideration to the methodological resources that are needed to generate deeper insights into this digital milieu, and in fact, we argue here that turning our attention to databases in particular could provide a wealth of insights into smart cities' social and political implications. To do so would build on, rather than negate, the important contributions of digital (and virtual) ethnography over the past two decades.

Specifically, we make three complementary arguments: that attention to the database can reveal the ways it, first, limits data types; second, categorizes and classifies data to align with its technical specifications; third, reflects the database designer's episteme; and fourth, (re)produces conceptions of the world. In each of these processes, knowledge about the smart city is delimited, framed, represented, and performed. In other words, open data serve as an important epistemological technology for smart cities. We substantiated these claims through an empirical exploration of Open Calgary, a key actant within Calgary's smart city movement. In sum, database ethnography can help researchers glean insights into social process, constituting an important counterpart to digital ethnography.

Acknowledgements:

Literature Cited

Atkinson P (2015) *For Ethnography*. SAGE Publications.

Atkinson P (2016) Digital ethnographies. *Qualitative Research* 16(2): 236–241. DOI: 10.1177/1468794114535041.

Author (Date) Details Withheld for Peer Review.

Barns S (2016) Mine your data: open data, digital strategies and entrepreneurial governance by code. *Urban Geography* 37(4): 554–571. DOI: 10.1080/02723638.2016.1139876.

Barratt MJ and Maddox A (2016) Active engagement with stigmatised communities through digital ethnography. *Qualitative Research* 16(6): 701–719. DOI: 10.1177/1468794116648766.

Beaulieu A (2004) Mediating Ethnography: Objectivity and the Making of Ethnographies of the Internet. *Social epistemology* 18(2–3): 139–163.

Beneito-Montagut R (2011) Ethnography Goes Online: Towards a User-centred Methodology to Research Interpersonal Communication on the Internet. *Qualitative Research* 11(6): 716–735.

Berger R (2015) Now I see it, now I don't: researcher's position and reflexivity in qualitative research. *Qualitative Research* 15(2): 219–234. DOI: 10.1177/1468794112468475.

Bietz MJ and Lee CP (2009) Collaboration in Metagenomics: Sequence Databases and the Organization of Scientific Work. In: *Proceedings of the 11th European Conference on Computer Supported Cooperative Work* (eds E Balka, L Ciolfi, C Simone, et al.), London, 2009, pp. 243–262. Springer-Verlag. DOI: 10.1007/978-1-84882-854-4_15.

Boellstorff T (2015) *Coming of age in Second Life: An anthropologist explores the virtually human*. Princeton University Press. Available at: https://books.google.ca/books?hl=en&lr=&id=jlLpCAAAQBAJ&oi=fnd&pg=PP1&dq=boellstorff+second+life&ots=wI1_nLytgi&sig=NNGv4Tm7rKDvDC3Qz0yY2S3O14Q.

Bowker GC (2000) Biodiversity datadiversity. *Social studies of science* 30(5): 643–683.

Bowker GC and Star SL (2000) *Sorting Things Out: Classification and its Consequences*. The MIT Press.

boyd danah and Crawford K (2012) Critical Questions for Big Data: Provocations for a cultural, technological, and scholarly phenomenon. *Information, Communication & Society* 15(5): 662–679. DOI: 10.1080/1369118X.2012.678878.

Brandusescu A and Sieber RE (2017) The spatial knowledge politics of crisis mapping for community development. *GeoJournal*: 1–16. DOI: 10.1007/s10708-017-9784-9.

Bratton B (2015) *The Stack: On Software and Sovereignty*. Cambridge, MA: MIT press.

Burawoy M (1998) The Extended Case Method. *Sociological Theory* 16(1): 4–33.

Canada NewWire (2011) Minister Day Launches Open Data Portal. *Canada NewsWire*, 17 March. Available at: http://ezproxy.lib.ucalgary.ca/login?url=https://search.proquest.com/docview/857414663?accountid=9838 (accessed 2 August 2017).

Collins HM (2001) Tacit Knowledge, Trust and the Q of Sapphire. *Social Studies of Science* 31(1): 71–85. DOI: 10.1177/030631201031001004.

Crampton J and Miller A (2017) Intervention Symposium: "Algorithmic Governance". In: *Antipode Foundation*. Available at: https://antipodefoundation.org/2017/05/19/algorithmic-governance/ (accessed 11 August 2017).

Dalton C and Thatcher J (2014) What Does a Critical Data Studies Look Like, and Why Do We Care? Seven Points for a Critical Approach to 'Big Data'. In: *Society and Space Open Site*. Available at: http://societyandspace.org/2014/05/12/what-does-a-critical-data-studies-look-like-and-why-do-we-care-craig-dalton-and-jim-thatcher/ (accessed 9 February 2017).

Dourish, P., and E. Gómez Cruz. 2018. Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society* 5 (2):2053951718784083.

Drucker J (2013) Performative Materiality and Theoretical Approaches to Interface. *DHQ: Digital Humanities Quarterly* 7(1).

Elwood S and Leszczynski A (2012) New Spatial Media, New Knowledge Politics. *Transactions of the Institute of British Geographers*. DOI: 0.1111/j.1475-5661.2012.00543.x.

Emerson R, Fretz R and Shaw L (2011) *Writing Ethnographic Fieldnotes*. 2nd ed. Chicago: University of Chicago Press.

England K (1994) Getting Personal: Reflexivity, Positionality, and Feminist Research. *The Professional Geographer* 46(1): 80–89.

Farnsworth J and Austrin T (2010) The ethnography of new media worlds? Following the case of global poker. *New Media & Society* 12(7): 1120–1136. DOI: 10.1177/1461444809355648.

Forte M (2004) Co-construction and Field Creation: Website Development as Both an Instrument and Relationship in Action Research. In: Buchanan E (ed.) *Readins in Virtual Research Ethics: Issues and Controversies*. Hershey, Pennsylvania: Information Science Publishing, pp. 219–245.

Gaffney C and Robertson C (2016) Smarter than Smart: Rio de Janeiro's Flawed Emergence as a Smart City. *Journal of Urban Technology*: 1–18. DOI: 10.1080/10630732.2015.1102423.

Geiger RS and Ribes D (2011) Trace Ethnography: Following Coordination through Documentary Practices. In: *System Sciences (HICSS), 2011 44th Hawaii International Conference on*, 2011, pp. 1–10. IEEE. Available at: http://www.davidribes.com/storage/Geiger%20Ribes%20hicss-trace%20ethnography.pdf (accessed 16 December 2016).

Gobo G (2008) *Doing Ethnography*. Thousand Oaks, CA: Sage Publications Ltd.

Graham M, Hjorth I and Lehdonvirta V (2017) Digital labour and development: impacts of global digital labour platforms and the gig economy on worker livelihoods. *Transfer: European Review of Labour and Research*. DOI: 10.1177/1024258916687250.

Gray J (2014) *Towards a Genealogy of Open Data*. SSRN Scholarly Paper, 3 September. Rochester, NY: Social Science Research Network. Available at: https://papers.ssrn.com/abstract=2605828 (accessed 27 April 2017).

Hammersley M (2006) Ethnography: problems and prospects. *Ethnography and Education* 1(1): 3–14. DOI: 10.1080/17457820500512697.

Harding S (1987) *Feminism and Methodology: Social Science Issues*, Bloomington: Indiana University Press.

Harley JB (1989) Deconstructing the Map. *Cartographica* 26(2): 1–20.

Herbert S (2000) For Ethnography. *Progress in Human Geography* 24(4): 550–568.

Hine C (2000) *Virtual Ethnography*. Thousand Oaks, CA: Sage Publications.

Hine C (2006) Databases as Scientific Instruments and Their Role in the Ordering of Scientific Work. *Social Studies of Science* 36(2): 269–298. DOI: 10.1177/0306312706054047.

Hinton S and Hjorth L (2013) *Understanding Social Media*. SAGE.

Hsu W (2017) A Performative Digital Ethnography: Data, Design, and Speculation. In: Hjorth L, Horst H, Galloway A, et al. (eds) *The Routledge Companion to Digital Ethnography*. New York: Routledge, pp. 40–50.

Kitchin R (2013) Four Critiques of Open Data Initiatives. In: *The Programmable City*. Available at: http://progcity.maynoothuniversity.ie/2013/11/four-critiques-of-open-data-initiatives/.

Kitchin R (2014) Big Data, New Epistemologies and Paradigm Shifts. *Big Data & Society* 1: 1–12.

Kitchin R, Maalsen S and McArdle G (2015) *The Praxis and Politics of Building Urban Dashboards*. The Programmable City.

Latour B (1986) Visualization and Cognition: Thinking with Eyes and Hands. *Knowledge and Society: Studies in the Sociology of Culture Past and Present* 6: 1–40.

Latour B (2000) Technology Is Society made Durable. In: Grint K (ed.) *Work and Society: A Reader*. Malden, MA: Blackwell Publishers Inc, p. 41.

--- (2004) How to Talk about the Body? The Normative Dimension of Science Studies. *Body & Society* 10 (2-3): 205-229.

Levy N (2016) Socrata cuts sales and marketing staff, shifts focus to large governments. Available at: http://www.geekwire.com/2016/socrata-layoffs/ (accessed 27 March 2017).

Luque-Ayala A and Marvin S (2015) Developing a critical understanding of smart urbanism? *Urban Studies*: 0042098015577319. DOI: 10.1177/0042098015577319.

Manovich L (1999) Database as Symbolic Form. *Convergence* 5(2): 80–99. DOI: 10.1177/135485659900500206.

Massey D (2003) Imagining the Field. In: Pryke M, Rose G, and Whatmore S (eds) *Using Social Theory: Thinking through Research*. Thousand Oaks, CA: SAGE Publications, pp. 71–88.

Mileham BLA (2007) Online infidelity in Internet chat rooms: an ethnographic exploration. *Computers in Human Behavior* 23(1): 11–31. DOI: 10.1016/j.chb.2004.03.033.

Miller D and Slater D (2001) *The Internet: An Ethnographic Approach*. London: Berg.

Murthy D (2008) Digital ethnography: An examination of the use of new technologies for social research. *Sociology* 42(5): 837–855.

Nakamura L (2002) *Cybertypes: Race, ethnicity, and Identity on the Internet*. London: Routledge.

Nast H (1994) Women in the Field: Critical Feminist Methodologies and Theoretical Perspectives. *The Professional Geographer* 46(1): 54-66.

Obama B (2013) Executive Order -- Making Open and Machine Readable the New Default for Government Information. Available at: https://obamawhitehouse.archives.gov/the-press-office/2013/05/09/executive-order-making-open-and-machine-readable-new-default-government- (accessed 3 August 2017).

Pinch T and Bijker W (1987) The Social Construction of Facts and Artifacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. In: Bijker W,

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

Rapport N (2000) The Narrative as Fieldwork Technique: Processual Ethnography for a World in Motion. In: Amit V (ed.) *Constructing the Field: Ethnographic Fieldwork in the Contemporary World*. New York: Routledge, pp. 71–95.

Robinson L and Schulz J (2009) New Avenues for Sociological Inquiry: Evolving Forms of Ethnographic Practice. *Sociology* 43(4): 685–698.

Schuurman N (2008) Database Ethnographies Using Social Science Methodologies to Enhance Data Analysis and Interpretation. *Geography Compass* 2(5): 1529–1548.

Sevigny D and Angelo L (2014) *Digital Strategy Report*. Calgary, AB: The City of Calgary. Available at: http://www.calgary.ca/cfod/it/Documents/CityofCalgaryDigitalStrategyReport.pdf (accessed 2 August 2017).

Shueh J (2016) What Obama Did for Tech: Transparency and Open Data. Available at: http://www.govtech.com/data/What-Obama-Did-for-Tech-Transparency-and-Open-Data.html (accessed 3 August 2017).

Sieber R and Johnson P (2015) Civic Open Data at a Crossroads: Dominant Models and Current Challenges. *Government Information Quarterly* 32(3): 308–315.

Simondon G (2017) *On the Mode of Existence of Technical Objects*. Minneapolis, MN: Univocal Publishing.

Srnicek N (2017) *Platform capitalism*. Malden, MA: Polity Press.

Star SL (1999) The Ethnography of Infrastructure. *American Behavioral Scientist* 43(3): 377–391. DOI: 10.1177/00027649921955326.

Star SL and Griesemer J (1989) Institutional ecology,translations' and boundary objects: Amateurs and professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social studies of science* 19(3): 387–420.

Thomer A and Twidale M (2014) How Databases Learn. In: 1 March 2014. iSchools. DOI: 10.9776/14409.

Townsend AM (2013) *Smart Cities: Big Data, Civic Hackers, and the Quest for a New Utopia*. New York: WW Norton & Company.

Tunçalp D and L. Lê P (2014) (Re)Locating boundaries: a systematic review of online ethnography. Linda Rouleau, Dr Mark de Rond and P (ed.) *Journal of Organizational Ethnography* 3(1): 59–79. DOI: 10.1108/JOE-11-2012-0048.

Wajcman J (1991) *Feminism confronts technology*. Pennsylvania State University Press.

Walker DM (2010) The Location of Digital Ethnography. *Cosmopolitan Civil Societies: An Interdisciplinary Journal* 2(3): 23–39. DOI: 10.5130/ccs.v2i3.1596.

Winner L (1985) Do Artifacts Have Politics? In: MacKenzie D and Wajcman J (eds) *The Social Shaping of Technology*. Philadelphia, PA: Open University Press, pp. 26–38.