

UCLA

InterActions: UCLA Journal of Education and Information Studies

Title

The datafication of transparency work: A report from Los Angeles. Proceedings for the Interactions Symposium on Big Data.

Permalink

<https://escholarship.org/uc/item/73j1q5sp>

Journal

InterActions: UCLA Journal of Education and Information Studies, 13(2)

Author

Currie, Morgan

Publication Date

2017

DOI

10.5070/D4132035205

Supplemental Material

<https://escholarship.org/uc/item/73j1q5sp#supplemental>

Copyright Information

Copyright 2017 by the author(s). All rights reserved unless otherwise indicated. Contact the author(s) for any necessary permissions. Learn more at <https://escholarship.org/terms>

Peer reviewed

In 2014 the City of Los Angeles' self-proclaimed "tech-savvy" mayor, Eric Garcetti used his third directive in office to announce the creation of an open data policy (Karlman, 2014). The policy requires each of Los Angeles' departments or bureaus, of which there are over 40, to select and release a dataset it manages on an open data portal, a public website where the data can be searched and downloaded in several machine-readable formats. Since the portal's creation, the City has also designed interactive maps based on its open data, showcased the City's progress on its service delivery goals on an open data dashboard, and hosted 'civic hackathons' that invite the public to make use of the data in phone app contests.

Garcetti's open data initiative is much more than a transparency effort; it is part of a larger policy goal to bring city government more in line with the technology industries that have begun to dominate twenty-first century economies. The policy aims for greater government accountability, but it also strives to unlock commercial potential through open formats that allow the data's reuse. For this reason, people are now likely to encounter Los Angeles' open data, whether they realize it or not, on popular commercial platforms. Yelp displays restaurant inspection scores from Los Angeles County's Sanitation department next to its restaurant details and reviews. The navigation app Waze uses the City's Bureau of Engineering data on road closures and construction permits. Google maps draws on the Metropolitan Transportation Authority's public transit schedules.

Under this policy, government records are more than an index of bureaucratic events; city records can also form the foundation for more data-driven service within government itself. Throughout my fieldwork in the offices of Los Angeles city departments,ⁱ city employees told me candidly that they hoped open data would revitalize departments and local economies, making the city more efficient and vibrant through data-driven policies and data sharing. Here is a quote from a government employee who anticipates open data will help city departments overcome the typical staid silos of bureaucracy:

I found myself wanting to go through every page of the new site. It's going to impact us internally. We're a city of work arounds. It's going to be transformative in this city because we'll be able to see what we're doing collectively [...] In terms of delivery of services, rather than an assembly line approach, we can have a bigger picture in mind.ⁱⁱ

In this proceeding I focus on how city employees such as this one reimagine their public records in light of open data. This reimagining occurs in at least two ways: first, as the basis for more data-driven policy, and second, as

financial assets available for commercial capture. Yet in order to bring about this new understanding, open data policy requires the *datafication of records*.

Translating Records as Data

Science and technology studies scholars Susan Leigh Star and James Griesemer (1989) have described how scientific data requires “translation” to move from laboratory observations into statistical charts, then to publishable research. Similarly, an amount of translation work goes on within city departments to convert city records to open data. Open data work translates records meant for specific departmental databases or spreadsheets into new statistical representations. This translation work has administrative consequences within City departments; Los Angeles hired a data coordinator to manage this work across several departments, and each department requires staff to do the grunt work of preparing and uploading data to the public portals.

This translation work entails that the traces of government activity become unmoored from the context of their original purpose. As a worker at Information and Technology Agency put it colorfully,

By definition, data is captured in such a way for a purpose, so each application is purpose-built . . . so when you’re making data open, you’re now identifying a new purpose. You brought a Rottweiler for home defense, and now you want an inside dog and need a transition to make everything work out okay.ⁱⁱⁱ

Open government data typically derives from departmental databases that describe actions taking place within that department as part of its mission, such as inspections filed and completed, permits approved, financial transactions made, even stray animals processed. When city staff generate a record, its creator in many cases has no concept of it having a broader external audience outside of occasional open records requests. Open data infrastructure, then, must do the work of translating closed-door bureaucratic practices to a wide, anonymous public.

This translation work involves what Alain Desrosières (2002) calls the creation of “spaces of equivalence” where large numbers of disparate events are recorded and summarized according to standard norms. All statistical work requires this kind of standardization in order to allow actions such as machine processing and statistical comparison.

In the case of open government data, staff must produce a second space of equivalence; that is, they must extract their records from bespoke databases, spreadsheets, and documents and make them compatible with *other* departments’ records by bringing them together into a common interface – the open data portal, a database of databases. Every dataset must be made available on the same

platform in machine-readable, open, and commonly used formats, such as .csv, json, rdf, and xml.

Through this translation work, open data loses the details of localized knowledge. Because government records are originally constructed without a broad public audience in mind, the particulars of their definitions and codes, their processes of recording and tabulating, often remain tacit institutional knowledge. Here is a quote from a systems analyst at Building and Safety,

We're a very technical organization, so most of the time that data has been formatted so that our engineers and inspectors could use it. I mean, you try to translate that to something that the public would understand; there's a gap.^{iv}

When records become data in open, public formats, there is a material transformation as data are formatted for the new platform. This transformation also entails a change in epistemic power. The power of *records* is indexical and evidentiary – a record is official evidence of a government action, grounding legal and political claims about phenomena. The power of *open data*, on the other hand, is to circulate and function through public reuse. The context of the record's creation as part of past government activities becomes less a matter of concern for open data's reuse than its circulation into other platforms, such as the Waze app or Google maps. The record, which is by standard definition a testimony of past actions (Trace, 2002), as *data* can be instrumentalized such that its role as evidence is less important than the ability to perform technical functions.

Take, for example, the Department of Transportation's (DoT) biggest open dataset, the traffic summary count. The count is gathered either by in-ground sensors or manually by DoT employees who stand for approximately six hours at an intersection and tally cars with a traffic counter device. Counts are not done at regular intervals but are usually conducted when commissioned through internal requests by other departments, such as Engineering, or by a commercial developer. A traffic-counting software package obtains the data from the counting devices and converts the data to PDF documents. Elements of the PDFs are also tabulated on individual lines in a summary excel spreadsheet. Both the PDF and the summary spreadsheet are available for download from DoT's website.

To contribute to the open data website, a DoT employee uploads the summary spreadsheet of traffic counts to the open data portal. On the portal, the user can make the traffic count perform several functions: it can be visualized as bar or pie charts or on a timeline; it can be downloaded as static data in CSV, JSON, RDF, or XML formats. A user can also access the data via the portal's application programming interface (API). By using the API, another platform such as a phone app or website can incorporate the data into its architecture. The value of the traffic count as open data lies in the technical affordances that increase its future, unanticipated uses.

Here is another example: The Department of Cultural Affairs (DCA) supplied to the portal one of its few existing sources of electronic information, their festival and events dataset. This dataset is a collection of 500–600 year-round events with date, location, and description. The information is typically accessed in a pdf form, but staff had created a spreadsheet of the information to publish on a now-defunct Wordpress site. To prepare it for the portal, a DCA employee extracted website URLs from description columns to put in a separate URL column and geocoded all festival addresses into the open data website’s acceptable geospatial formats. She described the process as fairly time consuming.

Once on the portal the festival and events information becomes machine-readable data. Like all datasets on the portal, the festival and events dataset can be visualized or ingested into other platforms through the API; the data can also be visualized as points on a map using the geolocation coordinates. Through the URL column, the dataset now becomes an access point to each event’s website. The DCA’s machine-readable data performs functions it could not as a pdf.



Figure 1. The front cover and a page out of the DCA’s festival guide, 2016.

event_name	event_link	event_description	start_date_and_time	end_date_and_time
1 Rose Parade	http://www.tournamentofroses.com	This historic and traditional California celebration featu	01/01/2014 08:00:00 AM	01/01/2014 11:00:00
2 Downtown on Ice Outdoor Skating Rink	http://www.pershingsquareicerink.com	Situated amid the Downtown Los Angeles skyscrapers,	01/01/2014 12:00:00 AM	01/20/2014 12:00:00
3 1st Thursday San Pedro Art Walk	http://www.1stthursday.com	On the first Thursday of each month, the artists, entert	01/02/2014 05:00:00 PM	01/02/2014 09:00:00
4 Undiscovered Chinatown Tour	http://www.chinatownla.com	Tours take visitors to a number of off-the-beaten-track	01/04/2014 10:30:00 AM	01/04/2014 01:00:00
5 Oshogatsu Family Festival	http://www.janm.org	Ring in the New Year and the Year of the Horse with fur	01/05/2014 11:00:00 AM	01/05/2014 05:00:00
6 Three Wise Men Festival	http://www.cervantescenter.org	A celebration of the Day of the Wise Men, a Hispanic tra	01/06/2014 12:00:00 PM	01/06/2014 03:00:00
7 Los Tres Reyes - The Three Kings	http://www.calleolvera.com	A celebration of the Epiphany (visit of the Three Kings)	01/06/2014 06:30:00 PM	01/06/2014 08:30:00
8 Australia Week 2014 / G'Day USA	http://www.australia-week.com	Australia Week is the largest foreign country promotion	01/08/2014 12:00:00 AM	01/14/2014 12:00:00
9 Downtown Art Walk	http://www.downtownartwalk.org	The Downtown Art Walk is a monthly, self-guided tour c	01/09/2014 12:00:00 PM	01/09/2014 10:00:00
10 Pomona Art Walk	http://www.metropomona.com	The Downtown Pomona Arts Colony has an art walk eve	01/25/2014 03:00:00 PM	01/25/2014 11:00:00

Figure 2. The LA Festival Guide as open data.

The Bureau of Engineering (BoE) offers a third example. Compared to DoT and Cultural Affairs, the BoE is a large, resource-rich, technocratic haven where their electronic systems are already set up for internal *and* public use. BoE already have staff devoted to maintaining and circulating electronic records internally and to the public. The staff I spoke with were enthusiastic about open data – the department’s principal civil engineer told me, “I was supportive of it, because we’ve always had that general mindset anyway.”^v Public data sharing was already a default mode of the department, which had begun building customized web software for public access back in the mid-90s.

Unlike the DCA’s and DoT’s data, which derive primarily from spreadsheets, BoE’s open data production entailed extracting data from its GIS land-based system. The datasets include all addresses in the city, all official street names, an accounting of public infrastructure, and political boundaries, such as council districts. BoE already makes this data public on its website Navigate LA, where the datasets are overlaid on a map interface. While the BoE data is publically available – already ‘open’ – the data cannot be easily extracted from the website in machine-readable form. The functions of that website are predetermined; they focus on access to the documents that underlie the data, such as providing hyperlinks to original permits and inspection reports at particular addresses. If a person wants the ESRI data itself, she would need to submit an open records request to receive it.

On the open data portal, in contrast, the BoE data are no longer overlaid but exist as discreet, downloadable sets. The portal allows the data to circulate

into other contexts, such that its openness is not only a matter of public access to content, but also one of formats and affordances that increase its potential for future use. The power of the record as open data lies as much in its *form* as its *content*.

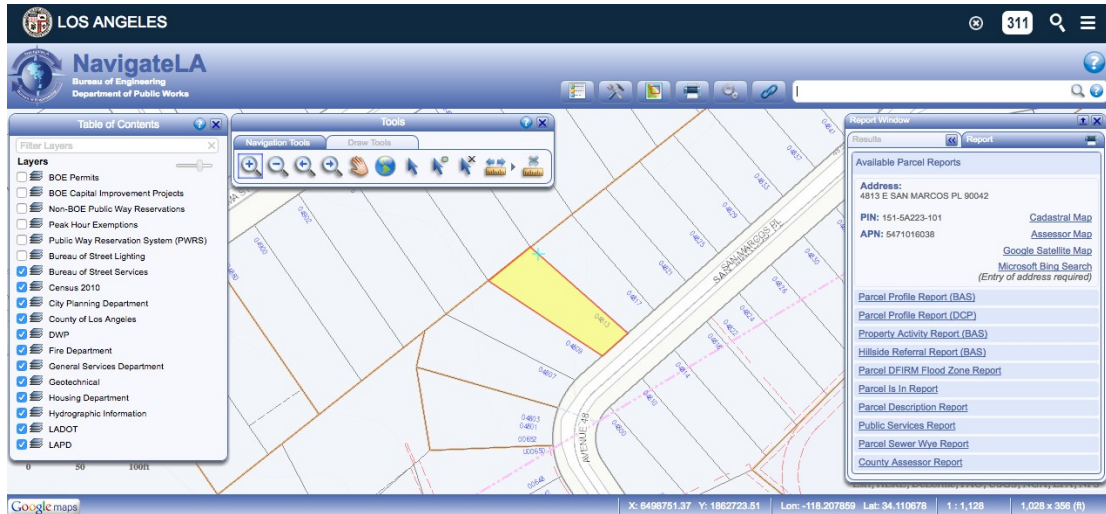


Figure 3. NavigateLA. The Table of Contents contains 527 layers of data. The Report Window allows users to access PDF documents on flood zone information, property activity, and county assessor records, among others.

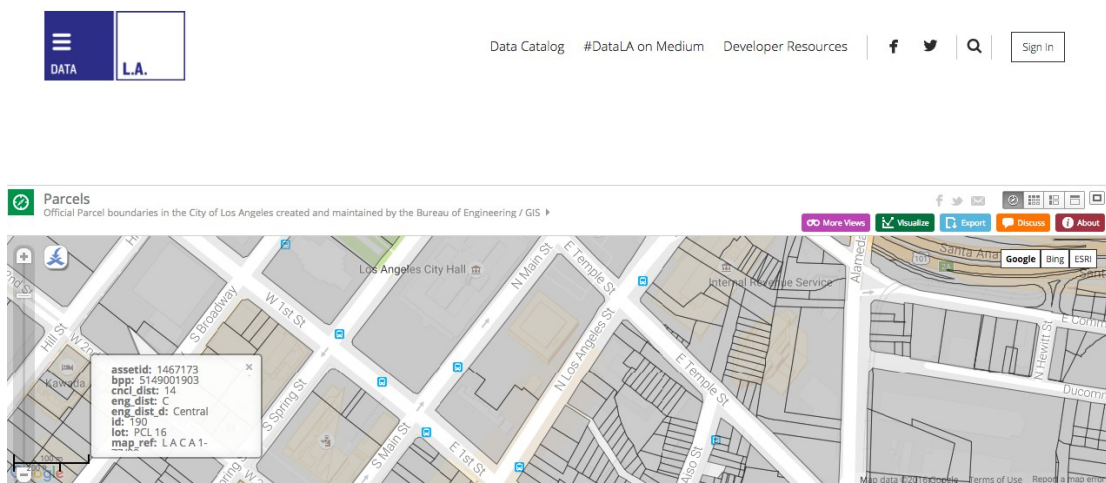


Figure 4. Bureau of Engineering's open parcel data. By clicking within the parcel lines a user can pull up identifying and geographic metadata for each parcel.

To be clear, in most cases City records become open data *not* because of some imagined public utility but because of the current state of legacy records infrastructure and the degrees of translation already in place to circulate records to a broader public. The DoT would have liked to contribute its data on street sign history, or work orders for maintenance activity on traffic signs and traffic signals, but did not because these records are still paper-based and their digitization would require significant labor and time. The same went for Cultural Affairs. Here's what one DCA employee told me:

We don't even have time to digitize our data. Everything we have is in paper form. We don't know what to do. Some of us don't even know what structure the filing cabinets and digital file directories on the intranet are designed to store our records. There's no time to get back into our filing cabinets. We're just pulling records as needed.^{vi}

In rarer cases, however, records were formatted as open data because the Mayor's office anticipated a new context for them beyond the portal. These contexts reflect how city employees have come to reimagine their records beyond their original purpose, even beyond their role as evidence of government processes.

In the next sections I focus on two ways that staff reimagine the role of City records through datafication: first, as an impetus for a public, data-driven approach to city management and second, as an information subsidy for commercial platforms.

Depoliticized service delivery

Even as most departments provide records that are simply easy to translate into open data, the Mayor's office asks departments to prioritize those records that could help with administrative decision making and tracking "progress towards key priorities" (Office of Los Angeles Mayor Eric Garcetti, 2013). When possible, departments supply the portal with performance metrics that populate the Mayor's dashboard, a website that keeps track of city goals such as homeless veterans housed, unemployment, graffiti removal, and pothole repairs.

Performance metrics have a history prior to open data policies. As a management style, these metrics became widespread after the 1992 bestseller *Reinventing Government: How the Entrepreneurial Spirit Is Transforming the Public Sector* by Ted Gaebler, a city manager, and David Osborne, a journalist (1993). Performance-based management departed from the standard public administration model that relied on pre-set regulations rather than outcomes and rewards for reaching administrative goals. Departments were to set these metrics themselves to establish a routine of indirect control by upper management. New Public Management (NPM), as this trend was called, asked departments to keep

track of their services to the public through statistical measures that could be analyzed in relation to outcomes and budgets. Whereas NPM was not overtly practiced in Los Angeles before the time of this research, open data revitalized aspects of this managerial approach.

Open data, however, puts a new spin on this management crusade. The Mayor's open data dashboard aggregates and publicizes the data, anticipating a data-literate public that is receptive to this form of representation and accounting. There the public can quickly survey the City's accomplishments or missteps. Data displayed in red means the numbers fall short; those in green hit the mark. At the time of this writing, the unemployment rate is looking good, at 4.6 percent; homelessness for veterans, on the other hand, is not at the desired goal of zero, with 172 still looking for homes.

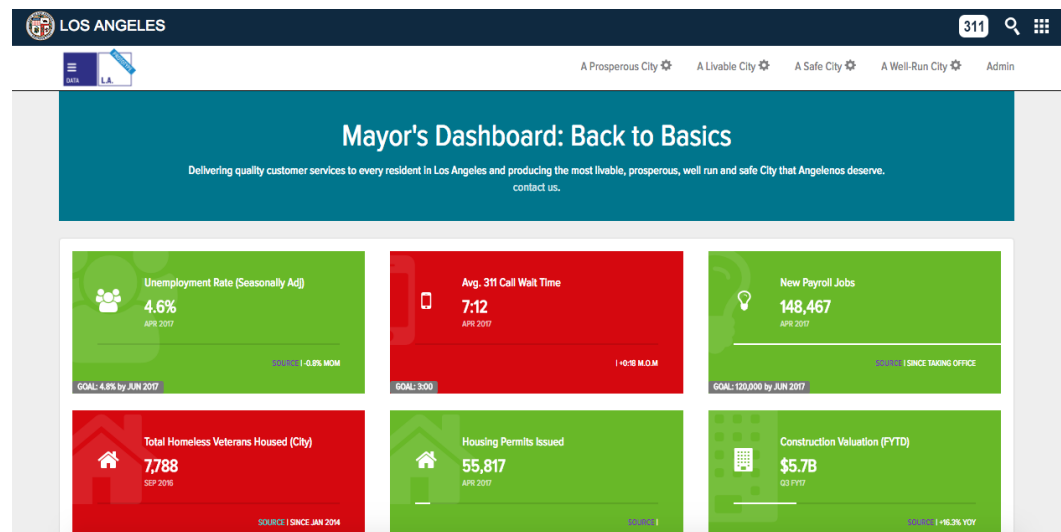


Figure 5. The Los Angeles Mayor's Dashboard.

Upon close scrutiny, the wider effectiveness of this interface is up for debate. The site provides limited context explaining the origins of the data, beyond hyperlinks to the dataset on the portal, even as the dashboard frames an argument about the datasets' importance. While appearing as authoritative metrics, none of the metrics are updated in real time or with any uniformity. And even as it purports to provide a partial window for citizens into government processes, the site offers no formal mechanism for participation into the trenches of administrative service delivery.

Furthermore, in most cases, the open performance data remains a passive snapshot of department activity, not a driver of departmental change. Said my DoT contact:

They don't offer any resources or any budgetary support to buy equipment, or software, or consulting hours, or whatever it would take to do these things . . . We don't even have enough people, and we don't, we're not able to fill our vacancies, so unfortunately, we're not in a good position to improve some of these things.^{vii}

One exception is the Department of Sanitation's Cleanstat data. In this case, the Mayor's office assisted the department with staff who designed an interactive map that draws on open data to show progress of the city's responses to illegal dumping, litter, weeds, and bulky items and to prioritize cleanup across 114 neighborhoods. According to a departmental survey, Cleanstat has made bulky trash pick-up more efficient, particularly in lower-income neighborhoods. While early 2016 saw 370 miles of streets requiring immediate clean-up, primarily of impassable sidewalks in the neighborhoods of South Los Angeles, by the end of the year the number had fallen to 91 (Poston, 2017). Cleanstat is one of the few clear examples of data-driven decision-making implemented since the portal's launch.

Open data appeals to proponents as objective measures divorced from the sway of political motivations, conscious or not. That open data be neutral is particularly important to government staff. According to an employee in the mayor's office, open data creates a neutral ground for consensus formation with the public:

A lot of times political problems can be challenged and questioned... If we do data-driven decision-making, we can take the politics out of decision-making and be more responsive to the public good and public demands versus if we don't.^{viii}

Open government data, then, remains tied to the common perception that data will offer a neutral guide for city problem solving and resource allocation.

The danger of more data-driven policy is that administrative decision-making will perceive the city as a technical problem to solve, "as a mere aggregate of variables that can be measured and 'optimized' to produce an efficient or normative system." (Mattern, 2015) Another concern is that administrative decisions will be made only on the basis of statistical evidence, rather than with other types of administrative knowledge, such as an experiential understanding of the widely varying types of communities being served in the metropolis. Data-driven services could be founded on – and encourage – an overly reductive view of urban phenomena.

Commercial capture

Proponents of open data in Los Angeles also spoke of data as a “raw resource,” an at-hand asset from which new value can be extracted at little cost. According to this industrial metaphor, once extracted, open data circulates within an ecosystem of reuse and profit, ready to be transformed by citizen innovators or private markets. According to an employee in the mayor’s office,

It’s a lifecycle to the data itself, in terms of it’s something that’s highly valuable, then tapers down after use . . . to a plain old toothbrush/toothpaste commodity, instead of something that produces increased efficiency. Something to figure out is: What’s the highest value in-house to look at?^x

A general manager at the Information and Technology Agency described this process in a statement that conflates the private sector and civic value of passive resources:

The private sector could be delivering to the public what we choose not to do or can’t . . . This is a different way of looking at government. It’s modern; it’s web 2.0. This isn’t our data; this is the citizens’ data. So long as it’s not inappropriate, the citizens and vendors know what to do with it more than we could. It’s very free market.^x

Mayor Garcetti announced an example of this free market in April 2015, during his second annual State of the City address (Office of Los Angeles Mayor Eric Garcetti). His office had made a data-sharing agreement with Waze that would affect the company’s 1.3 million local users. As a result, Los Angeles Waze users now receive data from the California Highway Patrol and Department of Public Works; the datasets identify road closures, film shoots, accidents, hit-and-run incidents, and AMBER alerts.

The data that provides Waze users with timely pop-ups on construction projects and road closures specifically derives from records created by the BoE. BoE is responsible for approving and managing construction permits within the public right of way. To process these requests, the agency uses a web-based electronic submissions system. Staff format these City records as open data so that they can circulate outside of BoE’s internal permit database and into Waze’s private interface, accessed by thousands of commuters in real-time.

In his announcement, the mayor tied the Waze partnership to the City’s open data initiative: “This is exactly what our open data initiative is all about – going beyond information sharing to provide opportunities for our data to transform into tools to make people’s lives safer and easier in very tangible ways” (ibid, n.p.).

Data sharing platforms that combine digital public resources with citizen-

generated and commercially-owned data blur the distinction between public and private information goods. I call these technical assemblages *data hybrids*. Public-private data hybrids provoke us to reexamine the classical definition of digital public goods, especially as these hybrid data infrastructures can, at times, encourage a form of corporate subsidy that makes private enterprise indistinguishable from the broader social good. Even more disturbingly, these hybrids create profitable flows that entangle the role of citizens with that of consumers who exchange their data for digital services. The partnership entailed no monetary transfer (as public goods the data on the portal are available for any type of use), and the data does not flow the other way, as a gift from private sector to public. Meanwhile Waze's parent company, Google, sees a financial profit by luring customers with a app that amalgamates user-generated, public, and commercial data.

Conclusion

The datafication of records changes the way we see our government and our cities. Open data enrolls government records in metaphors of data analytics and recasts public administration in the image of "big data" regimes. These policies do not straightforwardly reveal government processes and records; in the process of producing new statistical representations for the public, open data work subtly transforms administrative culture.

Other scholars have already pointed out that the concept of "openness" itself changes as open data policies are guided by this depoliticized and technical understanding of transparency (Yu and Robinson, 2012; Bates, 2012; Tkacz, 2013; Kitchin, 2013) Here I focused specifically on how, through open data work, staff view their records as administrative and commercial assets and fonts of innovation that improve private and public sector services. Some employees see city records as a source of easy capital, whether cost-saving efficiencies internally or innovation by the private sector. Others see open data as a solution for managing outsized, complex city problems.

Open data answers Los Angeles' need for new capital resources in a post-recession context and addresses a self-conscious concern about the lagging technological modernization of public institutions. This understanding in turn promotes data-centric management strategies and encourages a mechanistic understanding of how to solve city problems; it leads to new data hybrids born out of data-sharing partnerships between government and the private sector. When evaluating these new data infrastructures, the question is less one of whether the data is good or bad, since its indexical quality is more and more beside the point. Rather the data is valuable for how it performs and proliferates.

Notes

I. From November 2013 to August 2016 I spoke to government employees from nine city departments in the City of Los Angeles about the largely invisible clerical and technical work that goes into maintaining Los Angeles's open government datasets. I asked how each department prepares internal data for public consumption and reuse on the open data portal and about the datasets' departmental history. I also inquired about the cleaning, technical encoding, and maintenance required to create new datasets from existing electronic record infrastructures.

II. Interview conducted February 8, 2016.

III. Interview conducted July 21, 2015.

IV. Interview conducted August 21, 2014.

V. Interview conducted August 15, 2014.

VI. Interview conducted February 22, 2016.

VII. Interview conducted August 20, 2014.

VIII. Interview conducted February 8, 2016.

IX. Interview conducted February 27, 2014.

X. Interview conducted July 21, 2015.

References

- 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. *The Journal of Community Informatics*, 8(2). Retrieved from <http://www.ci-journal.net/index.php/ciej/article/view/845>
- Desrosières, A. (2002). *The Politics of Large Numbers: A History of Statistical Reasoning*. Harvard University Press.
- Karlamangla, S. (2014, November 22). Garcetti's high-tech, innovation agenda faces old-school hurdles. Retrieved October 27, 2016, from <http://www.latimes.com/local/california/la-me-adv-innovation-fund-20141123-story.html>
- Kitchin, R. (2013, November 9). Four critiques of open data initiatives. Retrieved October 30, 2016, from <http://progcity.maynoothuniversity.ie/2013/11/four-critiques-of-open-data-initiatives/>
- Mattern, S. (2015). Mission Control: A History of the Urban Dashboard. *Places Journal*. <https://doi.org/10.22269/150309>
- Office of Los Angeles Mayor Eric Garcetti. (2015, April 21). Mayor Garcetti Details Agreement with WAZE to Help Reduce Congestion, Increase Safety, and Improve Driving Experience Around L.A. [Text]. Retrieved June 9, 2017, from <https://www.lamayor.org/mayor-garcetti-details-agreement-waze-help-reduce-congestion-increase-safety-and-improve-driving>
- Osborne, D., & Gaebler, T. (1993). *Reinventing Government: How the Entrepreneurial Spirit is Transforming the Public Sector*. Plume.
- Poston, B. (2017, February 26). Garcetti promised to crunch data to clean L.A. streets — here's how that's working. *Los Angeles Times*. Retrieved from <http://www.latimes.com/local/lanow/la-me-ln-garcetti-back-to-basics-20170224-story.html>
- Star, S. L., & Griesemer, J. R. (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. <https://doi.org/10.1177/030631289019003001>

- Tkacz, N. (2013, January 28). Open sesame. Retrieved July 23, 2015, from <http://aeon.co/magazine/society/nathaniel-tkacz-open-source-government/>
- Trace, C. B. (2002). What is recorded is never simply “what happened”: Record keeping in modern organizational culture. *Archival Science*, 2(1–2), 137–159. <https://doi.org/10.1007/BF02435634>
- Yu, H., & Robinson, D. G. (2012). *The New Ambiguity of “Open Government”* (SSRN Scholarly Paper No. ID 2012489). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=2012489>