CHAPTER 4

Challenges to the Access of Government Open Data by Private Sector Companies

Peter A. Johnson and Christine Varga

Abstract

Many governments around the world are releasing open data, yet an understanding of how diverse stakeholders access this data is only just emerging. To understand how the private sector accesses and uses open data, interviews were conducted with Canadian information technology (IT) companies in the Kitchener-Waterloo to Toronto, Canada, high-tech corridor. Questions regarding how open data is accessed and used reveal what "access" to open data means for the private sector—seamless access across jurisdictions, access to a full catalogue of data, and access to accurate and current data. For governments that deliver open data, this nuanced reading of "access" can provide key feedback to improve current open data programs, and conceptualize the future of open data provision as an "ecosystem" of roles that governments could potentially fill, including as data creator, custodian, and provider.

Acknowledgements

We would like to thank the Geothink.ca partnership grant, led by Dr. Renee Sieber, for supporting this research. Funding for Geothink was provided by the Social Sciences and Humanities

Research Council of Canada (SSHRC). We would also like to thank Kevin Tuer, Communitech, and the Canadian Open Data Exchange (codx.ca) for facilitating connections with research participants.

Open data, though a relatively new concept, has rapidly become a dominant topic in the fields of IT, civic technology, and government information (Gurstein, 2011; Janssen et al., 2012; Sieber & Johnson, 2015). Typically provided by governments at all levels (municipal, provincial/state, federal), open data are made freely available through online portals, in machine-readable formats, and are shared under terms of a generous usage licence (Sieber & Johnson, 2015). Open data is considered to be one of the key ways in which governments can deliver on the transparency and collaboration principles of open government plans or strategies (Ruvalcaba-Gomez et al., 2018). As the provision of open data expands, questions about its value and use are posed, particularly concerning the value of open data to specific user communities (Janssen & Zuiderwijk, 2014; Johnson, 2016). The private sector—that is, profit-maximizing companies, that are either from sole proprietorships or to publicly traded entities—is frequently identified as one of the major user communities for government open data (Deloitte, 2012a; Bonina, 2013; Ruppert, 2015). Despite this, there is little direct research on how the private sector interacts with government open data, and what barriers may exist to access. As part of the rapidly developing open data "ecosystem" (Heimstädt et al., 2014; Sangiambut & Sieber, 2017), it is important to address the differential needs and preferences of diverse user communities. This research aims to fill this gap, using interviews with key stakeholders at a variety of private-sector companies in the Waterloo–Toronto, Canada, IT corridor to better understand how open data, typically provided by municipal governments, are accessed and used, identifying key challenges that restrict this use. We use these empirical findings to frame a discussion of strategies that government open data providers can employ to develop an open data ecosystem that is more responsive to the needs of the private sector. In this instance, the future of open data is one where data-producing governments can better connect their data with the specific needs of an identified non-government user base. The simple provision of open data was an early challenge, and moving forward, the future of open data should begin to tackle the challenge of facilitating use.

We conclude with a discussion of the roles that governments can play within this open data ecosystem, including as data creator, data custodian, and data provider, and the potential opening up of these roles to non-government actors.

1. Measuring the Use and Value of Open Data

It has traditionally been challenging to track how users access and work with open data. The very nature of open data as free from access restrictions can make it difficult for open data provisioning governments to gather metrics of use (Johnson, 2016; Johnson & Greene, 2017). When considering quantitative approaches to measuring the value of open data for private sector use, studies note that value cannot be determined through correlations with factors regarding the provision of, or access to, the data. Fumega (2014, p. 29) refers to attempts to count the number of website visits, published datasets, or downloads as "flimsy metrics," and notes that conclusions based on these measures are unlikely to be robust. Harrison, Pardo, Cresswell, & Cook et al. (2011) explain that metrics which attempt to quantify the number of datasets or the opportunities for participation and collaboration do not always indicate value. Similarly, Deloitte (2012a) notes that the number of downloads, or "clicks," on a dataset cannot be directly equated to economic benefit. They explain that when using the number of downloads as a proxy for demand, "detailed quantitative estimates of economic impact can then only be established if such demand can be positively correlated and causally linked to conventional measures of economic output per sector" (Deloitte, 2012a, p. 8). These estimates cannot always be established, however, due to the challenges of tracing use of open data once they are downloaded from a government open data portal. Simply put, there are few tools that governments have at their disposal to trace the use and impact of open data by a broad range of end users (Johnson, 2017).

Existing studies have highlighted a variety of challenges to the use of open data by the private sector. For example, in a study by the Open Data 500, an international network of organizations that studies the use and impact of open data, the most significant challenges found for users of open data were access, accuracy, and level of detail (Australian Government, 2015). In Fumega's (2014) case studies, common obstacles for open data users included overemphasis on technical aspects and lack of usability for non-technical users. Other issues

found with the data include lack of standardization, information that is not up to date, and too much "noise" in the data (Latif et al., 2009; Manyika et al., 2013). Zuiderwijk, Janssen, Choenni, Meijer, and Alibaks et al. (2012) take analysis of these and other issues further, by identifying 118 impediments to open data use. Their findings are divided into 10 impediment categories: availability and access, findability, usability, understandability, quality, linking and combining data, comparability and compatibility, metadata, interaction with the data provider, and opening and uploading. Not only are there many possible issues with the data, but these issues are proven to be present in many cases. In the study by Sayogo et al. (2014), it was found that only 66% of the existing open data portals provide the ability to manipulate the data, and that only 49% of the existing data portals provide data in formats that support linked data. These statistics, as well as the long list of potential obstacles to reuse of the data, demonstrate major areas for improvement in easily accessing the data and using it to capture value.

Davies et al. (2013) see two possible future paths for open data impact research. The first is analysis at the macro-level, observing statistical correlations between open data implementation efforts and outcomes that imply some expected impacts of open data. They give the relationship between economic growth and levels of open data publication as an example, with governments that publish large open data catalogues spurring economic growth and innovation, typically in the IT sector. Jetzek, Avital, and Bjørn-Andersen et al. (2013), have conducted this type of macro-level analysis, searching for correlations between four enabling factors (exogenous variables) and four value-generation mechanisms (endogenous variables). The exogenous variables are openness, data governance, capabilities, and technical connectivity, while the endogenous variables are efficiency, innovation, participation, and transparency. As a result of the analysis, all relationships between exogenous and endogenous variables were supported except for the openness-to-transparency relationship, for which a slight negative correlation was found. Jetzek et al. (2013) find results that provide some merit for their efforts, yet it should be noted that these correlations encompass a broad definition of value that is not limited to the private sector.

A common quantitative approach to measuring the value of open data is the attempt to estimate value in terms of currency (Carrara et al., 2015; Gruen et al., 2014). Two of the most extensive and most commonly cited efforts to estimate the economic impact of open

data are a 2011 report from the European Commission, and a 2013 report by McKinsey & Company (Manyika et al. 2013). These and similar studies and other similar efforts review previous work on estimating open data's value, or explore research from multiple sectors and attempt to aggregate the findings. The European Commission's paper report (see Vickery, 2011) concludes that if current public sector information was available for free or at marginal cost, data-use activities could increase by up to €40 billion each year, as compared to the case where the data is not open (Vickery, 2011). McKinsey & Company find $3 trillion in annual economic potential globally through the release of open data. This value is not the result of an extensive study of all sectors (Manyika et al., 2013). Similar to efforts to determine value through correlations, these estimates of value are not limited to the private sector. Despite these findings produced by this quantitative approach to measuring the impact of open data, there is a level of nuance missing. Rather than an instrumentalist focus on connecting data provision to outcome via quantitative measures, we propose to build on Johnson and Greene's (2017) work that conducted qualitative interviews with public sector open data providers to better understand their process of data provision and use tracking. This qualitative approach matches with the second path proposed by Davies et al. (2013) for open data impact research, a micro-level analysis of the processes through which open data is used. This research places a focus on understanding how private sector start-up IT companies access and use open data, and the challenges to their access and use.

2. Interviews with Private Sector Open Data Users in the Kitchener/Waterloo–Toronto Corridor

To refine our understanding of how private sector companies access open data, individuals from 11 Ontario-based companies were interviewed, representing finance, logistics, mobile-app development, data provision, and IT consulting. These companies are based in the Kitchener-Waterloo to Toronto corridor, Canada's most significant cluster of IT companies. Interviewees were from a combination of start-ups and established companies, with study respondents recruited through connections with the Canadian Open Data Exchange (www.codx.ca), a public–private partnership agency founded to support the use of open data in private sector companies. The interviews took approximately 30 minutes, and were conducted

either in person or by telephone. Interviews were audio-recorded, transcribed, and coded for main themes, including benefits of open data and challenges faced in accessing and using open data. The results of these interviews are presented in two sections. The first is a descriptive section that details the classification of users, sources of open data accessed, and description of the benefits of using open data. The second section presents responses from the respondents, including on the challenges to using open data, including access to data across jurisdictions, access to the full data catalogue, and data accuracy and currency.

2.1 Classification of Users

Using the open data user classification developed by Deloitte (2012b), which divides private sector open data users into five categories, we found that respondents included three categories of users who employ open data to support their operations: aggregators, developers, and enrichers. Of the companies surveyed, six are providers of new products or applications built using open data. One company acts as an aggregator, providing processed open data to clients, while the other five companies are considered developers of new applications based on open data. There is overlap between these latter two categories of user, with some respondents displaying characteristics that point to both the aggregator and developer user definitions, given their diverse product lines. Unlike these new product and service providers, five individuals identified their organizations as enrichers, because of their use of open data as an input into existing products and services. As one respondent said, "the product is what people are interested in, not the fact that it's just a bit of open data." All three user types noted the need to generate a value-added component to their product, compared to simply replicating the raw data available directly from the government open data catalogue. In this way, the respondent pool was largely working with open data as a base material that would be used to create or combine with other data and services to form a saleable product.

2.2 Sources and Types of Open Data Accessed

Private sector open data users that were interviewed accessed open data from a wide variety of sources from Canada, the US, and around

the world. The specific sources of open data accessed were closely related to a particular project or client of the company. One respondent noted, "our big focus is in Canada and the US, primarily because that's where we are and that's where the market is ripest and richest." Open data sources varied based on the application area. Examples of open data sources given included government-owned or affiliated sources like Statistics Canada (www.statcan.gc.ca), openNASA (open.nasa.gov), World Bank (www.worldbank.org), Land Information Ontario (www.ontario.ca/page/land-information-ontario), Environment Canada (weather.gc.ca), and various provincial ministries or local municipalities, and other non-government associated open data sources like OpenStreetMap (www.openstreetmap.org). Some of the specific datasets being accessed by these companies include geospatial data, like flood data, terrain, aerial photography, land use, building footprint, building heights, zoning, surficial geology, groundwater data, forest cover, woodlots, and weather data. This is consistent with findings from other studies that note geographic or geospatial data to be the most heavily used open data in the private sector (Australian Government, 2015; Greene & Rinner, 2021 this volume). Other examples given include demographic sources like poverty indexes, macroeconomic indexes, population and housing projections, or financial data, business registry data, licence data, and government procurement information. These examples are consistent with findings from previous studies. At the local level, sources being assessed included transit data, traffic collision data, local points of interest, event information, and polling data. Given the wide range of open data topics available across many different levels of government, the respondent pool indicated that data sources and topics were selected on a case-by-case basis, dependent on the needs of a particular client or project.

2.3 The Benefits of Using Open Data

Past literature identifies methods of generating value through private-sector open data use, including combining open and proprietary data, integrating open data into a particular analysis, and the development of novel applications (Johnson, 2016; Australian Government, 2015; Granickas, 2013; Kassen, 2013; Manyika et al., 2013). By interviewing private-sector users of open data, we aimed to determine the benefits, challenges, and constraints to the use of open data. When

asked about the benefits of open data use for private sector companies, financial benefit was often mentioned first by respondents. Respondents also noted that the ability to charge for their products and services is based on the value-added customization provided in addition to, or "over top" of government open data. One company indicated that being able to access open data from various jurisdictions allows them to vend a similar product to many clients, explaining that "the idea would be, we can provide the application, scale it to as many clients as we would want and [our customers] would have a more straightforward means of populating our application with their data sources." In this sense, as open data becomes easier to access and use, the company will be able to support more clients in using open data within their analysis ecosystem.

For companies who identify as enrichers, open data was seen to add value to existing analytical products at minimal cost, allowing the company to receive some of the same data for free that it would otherwise typically need to purchase. The use of free open data effectively reduced costs to the company, and allowed for a more cost-competitive product to be offered to the end consumer. This also allowed companies to create prototype or "test" models and products more quickly, adding a free or low-cost tier to their product offerings allowing companies to test out a number of different strategies without needing to commit to expensive proprietary data sources. As one respondent noted, "it's a great way to start a business where your supply chain is a free product." In some cases, data these companies rely on are only provided by governments, thus the provision of these data as "free" is a significant way in which government and taxpayer funds were seen to support the expansion and profitability of the private sector.

Specific types of open datasets were seen as more valuable than others. Given the range of open data provided by many governments, that there are certain "key" datasets of interest to private companies is not surprising. As is consistent with the past literature, respondents noted that market-related data were most valuable (Granickas, 2013). As one respondent pointed out, "you can just imagine what it would do for the real-estate market to have access to business directories, locations of parks, things like that." Demographic data were also of interest, with one respondent noting, "governments tend to have the best demographic data. The only companies that can really compete with that are some of the very large retailers like Walmart and Google." Data from these companies are expensive to acquire, so open

data accessed at no cost to the users is often a better investment, despite its potential limitations.

Another benefit of open data identified by private sector open data users is the combination of multiple datasets, especially the aggregation of government data and other types of non-governmental data. This aggregation across scales of government and geography is widely identified as one of the potential benefits of a broad ecosystem of government open data (Janssen et al., 2012). One research respondent explained how this aggregation of data could be valuable to their organization, particularly when compared to proprietary data sources: "We have to have hundreds of sources merged together in order to have a compelling counter argument to the closed data that people are used to because it's comfortable." Many respondents noted that they always use more than one data source, and in some cases used up to 25 unique data sources for certain projects. Most respondents also noted that they engaged in some aggregation of public sector open data with private sector data from other sources. These sources could include data from marketing companies, data provided by clients, or data found by scraping other websites. One respondent also noted that they were seeing companies aggregate open data with data generated internally by the company itself, a process of "accenting big data with open data." Some companies are also engaged in the process of combining multiple open data sources across jurisdictions to derive insights that cross geographic areas. Across the private sector respondents interviewed, the value of this aggregation of open data and its combination with proprietary sources was seen as a key benefit of the provision of open data.

3. The Challenges of Using Open Data

Despite several benefits identified by respondents about the use of open data to support their business, there was significant commentary on the challenges of accessing and using open data, including data format, standards, quality, and technical constraints. Many of the respondents were also concerned with the relative lack of strategies to overcome these open data access and use challenges. This lack of existing knowledge on overcoming challenges is likely due to the emerging state of open data use by the private sector, as well as the lack of data provision standards between data-delivering governments. This section of results presents three key dimensions of

"access" identified by research respondents as challenging: access across jurisdictions, access to the full catalogue, and access to accurate and current data.

3.1 Access across Jurisdictions

Respondents frequently mentioned issues with access and availability of data across jurisdictions. For example, a significant issue was found with respondents needing to access many different government open data catalogues to piece together sufficient data to meet client or project requirements. This need to access data that crosses jurisdictions was further aggravated by the different approaches used by each government to collect and deliver similar data types. One respondent explained: "Federal information is not often comparable to provincial information . . . I'm stuck working at cross-purposes, where I'm kind of extracting bits and pieces of various datasets at the federal level and various datasets at the provincial level to get comparable data, and it gets even worse when I start looking country to country." Respondents noted that it is difficult to go to many different government catalogues to get all the data they need. This time-consuming searching for data was augmented by additional post-processing and data formatting required to create a blended dataset. Overall, the lack of seamless access across varying levels of government and jurisdictions led to fragmentation of projects and difficulty in developing business models at the required scale to be profitable.

3.2 Access to the Full Catalogue

Respondents often noted that not all data that businesses would like to access are actually available through an open data catalogue. One respondent explains, "I know there are datasets that I would like to be able to access, but there hasn't been a consistent and coordinated effort on the part of the government agencies to provide it." Another respondent echoed this, stating, "I'm not sure governments are completely open to releasing all the data that they should be releasing." Given the significant repositories of data collected by governments (particularly provincial or federal), making or prioritizing specific data releases can be challenging. Recent efforts on the part of governments to publish comprehensive lists or inventories of all data that could possibly be made open are a significant step towards improving on this issue of

access, facilitating more targeted data requests by end users. Similarly, government-open-by-default policies, where data produced are considered open when created, and only protected by specific request or through a privacy assessment, will also serve to reduce gaps in open data catalogues that can impede open data access.

3.3 Access to Accurate and Current Data

The accuracy of data obtained through open data portals was an issue impeding use by the private sector. Most respondents had noted some issues with datasets that may not be "ground truthed" or verified with actual locations or features. This issue stems from the high cost of checking datasets for quality before they are used in a paid product. One respondent explained that they would need to retreat to using paid data to be confident enough to use the data as part of a paid commercial product. Another respondent commented that since their company would need to take responsibility for inaccuracies, "a lack of data, in my opinion, is better and something I'll accept over wrong data." An opinion expressed by one respondent was that private sector companies need to demonstrate the value of the data to governments before they will commit to improving quality. They explained: "The city could go through the trouble of creating an open dataset, adhering to the standard and checking that set for completeness because they anticipate that it's going to be used in something, or they know it's being used in something that's returning value to them in some way, shape or form, then they are incented to make sure at least it's good." Some respondents gave an alternate opinion, claiming that governments must ensure data are accurate before they are published, and that this series of checks creates unnecessary delays in data release: "I think governments are good at making sure the numbers are correct, which is a really good thing that they are doing. The problem is I think that's actually slowing down update frequency, so there needs to be a better balance between the two." This lack of timeliness in publishing updates was a major issue noted by nearly all respondents.

Timeliness, or data currency, was mentioned as an issue by several respondents. In regard to demographic data, such as a census, a respondent explained: "Many countries do it once every five years. That is not sufficient for our clients. We still have to add commercial data to it, to make sure that it is updated every year." Nevertheless, most respondents agreed that improvements to future publishing

timelines would be a positive development. A respondent explained, "in a perfect world, you would have real-time data of everything, allowing an infinite number of use cases. Whatever your imagination is in terms of a way to extract value from that data, you're not limited because it is data that can be called through an API [application programming interface] and it is relatively real-time streaming of that data, so that you can do everything." Overall, given access to more current and accurate data, respondents felt that open data could be reliable as an integral component of new commercial applications.

4. Discussion: Improved Linking and Standardization of Datasets to Support Private Sector Use

The findings from this research indicate that there are select openings for governments to support the use of open data by the private sector. When considering the future of open data provision by governments, there is a strong case to be made in improving data access as a way to increase usage. Specifically, we aim to answer the question of how a government that provides open data can also support the access and use of these data with private sector users in mind. Moving forward with the maturation of government open data, there are key actions that can support more than the simple provision of open data (the "data over the wall" approach, according to Sieber & Johnson, 2015), creating an open data ecosystem (Heimstädt et al., 2014). This more fully developed open data ecosystem brings together data providers and end users to consider how to best deliver open data to satisfy the various needs of constituent end user communities. Through the results of our research with private sector open data users, we propose two core actions that governments can take in the near future to support the development of an open data ecosystem. These are: (1) development and support for linked open data, and (2) improved adoption of common open data standards whenever possible.

4.1 Linked Open Data

A first notable opportunity for improving open data published by the government is the development of linked open data. Linked open data are both human and machine readable, and adhere to uniform conventions of naming and linking, allowing data to be easily connected, queried, and even integrated. For example, by offering linked open

data, governments may be able to partially solve the jurisdictional variation issues, creating a link in metadata that allows similar datasets to be connected across jurisdictions. While a recent study finds that only four European countries officially support linked data (Carrara et al., 2015), it is also clear that linked data are an important element of data provision and functions to make data more accessible and usable. In discussing what he calls the "linked value chain," Granickas (2013) sees linked data as a cost saver, as information can be easily found, and connected, moving from data simply provided, to data embedded in a linked network of data sources, much like the Internet exists as a collection of connected documents. Ubaldi (2013) also notes that linked data are required for more sophisticated queries, particularly those that cross geographic scales, or connected themes.

Current work on developing linked open data builds on early development of the "semantic" (or machine-readable) web (Berners-Lee, 2009). Currently, the World Wide Web Consortium (W3C) Linked Open Data community project aims to develop a global data commons of open datasets linked through the use of resource description framework (RDF) links that make machine-readable connections between data. From a government perspective, as a data creator, custodian, and publisher, adding RDF metadata to a given dataset would enable these types of connections based on geography, topic, content, and other data characteristics (Ruback et al., 2016). Taking existing government data and creating linked open data is not necessarily a trivial step, necessitating changes in the way that data are created and shared. Additional resources, including staff time and an advanced knowledge base, are required for municipal staff to not only provide data, but to link them and maintain those links. The adoption of linked open data from a data provider perspective hinges on continued support of and investment in the broader role of government as a key open data provider. This future, though optimistic, is not a given, as governments continually must respond to varying levels of political will to invest in open government and transparency initiatives, as well as the often slow pace of government IT projects. Though potentially valuable as a contributor to private-sector use of open data, the broad adoption of linked open data is still ongoing, presenting a notable area of future attention for data-providing governments.

4.2 Greater Standardization of Datasets

Implementing and following data standards is a significant way that governments can support improvements to the open data provision ecosystem, particularly private sector users of open data. According to Davies et al. (2013), the development and enforcement of standards between different open datasets can simplify data sharing and use. Ensuring that data are produced and structured according to commonly held standards improves data coverage over many jurisdictions, and also allows for common analysis and development tools to be used. As revealed by the respondents, and mirrored in open data literature, from a private sector user perspective, a lack of data standardization leads to greater challenges in bringing together datasets from different jurisdictions, or replicating tools and analyses from one area in another area (Janssen et al., 2012). A lack of use of data standards can create a situation where data from one jurisdiction are not structured in the same way as data from another jurisdiction, necessitating multiple pre-processing steps to create a common frame for analysis and development. From a private sector perspective, this lack of standards between different datasets can prevent the generation of economies of scale that would be required to create profitable services (Zuiderwijk et al., 2012).

The implementation of data standards is a challenging process, requiring time, effort, resources, and coordination between different data producers and standard developers or proponents (Plu & Scharffe, 2012; Zuiderwijk et al., 2012). A handful of open data standards have been successfully adopted, most notably the General Transit Feed Specification (GTFS), which sets standards for how real-time transit data are stored and shared (McHugh, 2013). Given the additional value to be derived from increased standardization, not only for private sector users but for all open data users, the continued development, promotion, and adoption of standards are critical areas of ongoing work. Current work on developing an open data standards directory (see http://datastandards.directory) is an important step towards sharing information about available standards, a key precursor to enabling standard adoption. As referenced by our respondents, to truly support an innovation ecosystem and provide service to private sector users, government, in its role as data creator, custodian, and provider, should increasingly look to adopt existing open data standards, and work across government jurisdictions and

with the data user community to increase the use of standards in open data provision.

5. Conclusions

When considering the future of private sector use of open data, it is important to note the critical role that governments play as data creator, custodian, and provider. Governments may play all three of these roles or potentially only one or two, offloading specific roles to the private sector or other entities. How each of these roles that governments play can be managed, affects private sector use of open data, and the ability for open data to fulfill even partially the much-hyped "innovation" agenda that often drives data release (Bates, 2014; Zuiderwijk et al., 2014). For example, data creation roles involve adherence or non-adherence to standards (where they exist), data-custodian roles can restrict certain aspects of a dataset or entire datasets, depending on privacy and data quality concerns, and the data provider role can serve data to end users in a variety of formats and frequencies, which may or may not fill the needs and preferences of varied end user communities. When governments look to provide data through an open data platform, many factors come into play, including ease of publishing, perceived demand for a specific dataset, overall value of data, as well as other dataset-specific issues, such as quality and completeness. The results of this research support existing literature that shows the challenges of open data provision (Janssen et al., 2012; Sieber & Johnson, 2015; Johnson et al., 2017), but frames these challenges from a private sector use perspective, informing the current state of knowledge with their unique needs as an important user community. From the perspective of our respondents, there is a notable lack of consistency between government open data providers in what a data user could expect to find in an open data catalogue, both in terms of dataset availability and in the relative quality of any given dataset. These factors serve as a damper on the possible use of government open data in a private sector context, restricting the generation of value from the provision of government data. Additionally, there is risk that governments be removed or replaced as a provider of certain types of specialty data, as private sector organizations seek to overcome access challenges implemented by laggard governments (Johnson, 2019).

To rectify these shortcomings, our sample suggested several possible strategies. Of these, there are straightforward suggestions,

such as ensuring that data released are of the highest quality and completeness possible, and are released in a timely manner. Pushing governments, particularly at the municipal level, towards releasing a common suite of datasets that represent those most common or critical datasets, would help to provide better coverage across areas. This research supports existing work on understanding and promoting the concept of "linked open data" (Ubaldi, 2013), as a way to improve access to open data, creating a valuable part of an open data ecosystem. Linked open data are structured so that metadata provide links with other related datasets. From a private sector perspective, a linked dataset eases discoverability of other related datasets, saving time and money (Granickas, 2013; Janssen et al., 2017). Lastly, open data standards also have a strong role to play in supporting access to open data and enabling reuse, allowing users to scale up projects more quickly and seamlessly transfer work from one jurisdiction to another. Many of the respondents noted that they want governments to work towards these initiatives, including developing an open data ecosystem that crosses jurisdictions, linking open data, and increasing the standardization of datasets. Despite this interest from the private sector in having governments lead these challenges, it remains to be seen who may be driving these types of initiatives. An example would be the development of a third-party open data catalogue that crosses many jurisdictions, or federates a number of municipal catalogues into a higher-order regional or provincial data catalogue (Johnson, 2019; Wang & Shepherd, 2020). Similarly, there is potential for private business to assume the role of data enricher and re-seller, taking government data and repackaging them for other audiences. Given this encroachment on government open data provision, future work in this area needs to address not only the constraints to data access that may challenge a particular user community, but also the potentially changing role of government to that of the data custodian, abdicating the role of data provider to the private sector.

Despite the existing opportunities for governments to better meet the needs of private sector open data users, there remains a potential risk that in shifting from the simple provision of open data towards tailoring to the needs of one specific user community, governments may create areas of disadvantage for other user groups (Yang & Wu, 2021). For example, government data-provision resources may become exhausted through meeting the technical requests from

private sector or large-scale users, reducing government capacity to invest in the provision of other, less commercially relevant datasets (Johnson et al., 2017). Additionally, this shifting focus towards the private sector can also create an open data ecosystem where all users, including from civil society, or the public sector, would need to make a business case or demonstrate economic value for data to be opened or provisioned in a specific way. Though data-providing governments can better meet the needs of specific user groups, this should not come at the expense of service to other user groups, or the development of unequal or selective pathways for data access.

Given the potential for government priorities to change, particularly when it comes to open government policies and open data programs, there remains the possibility that data will cease to be open, disrupting private sector use of open data (Johnson et al., 2017). Though there is a tendency to view open data as a permanent resource that will continue to be provided by governments, there is a possibility for the existence of open data to contract rather than expand. Advocates of open data need to demonstrate returns in order to secure more support and financing. This "impermanence" of open data could be driven by government reaction to any number of local or global trends, as well as a further retrenchment of government as a direct service provider to citizens and the development of a consultative layer (Brabham & Guth, 2017) that enters to mediate connections between government and data end user. Government policies and actions change over time, and if open data is not deemed beneficial, it may cease to exist. This permanence, or "online stickiness," of data is a key foundation in governing open data (Sunlight Foundation, 2010). Absent the expectation that data are consistently available, use from all sectors, whether private, public, or not-for-profit, will be restricted. As such, open data becomes less "open" and less valuable if they are not permanent.

References

Australian Government. (2015). *Private sector use of open government data: Results from the Open Data 500 Australia.* https://www.communications.gov.au/publications/open-data-500-report

Bates, J. (2014). The strategic importance of information policy for the contemporary neoliberal state: The case of open government data in the United Kingdom. *Government Information Quarterly, 31*(3), 388–395.

Berners-Lee, T. (2009). *Linked data.* https://www.w3.org/DesignIssues/LinkedData.html

Bonina, C. M. (2013). *New business models and the value of open data: Definitions, challenges and opportunities.* Department of Management, London School of Economics and Political Science. http://www.nemode.ac.uk/wp-content/uploads/2013/11/Bonina-Opendata-Report-FINAL.pdf

Brabham, D. C., & Guth, K. L. (2017). The deliberative politics of the consultative layer: Participation hopes and communication as design values of civic tech founders. *Journal of Communication, 67*(4), 445–475.

Carrara, W., Fischer, S., & van Steenbergen, E. (2015). *Open data maturity in Europe 2015: Insights into the European state of play.* Capgemini Consulting. https://www.capgemini-consulting.com/open-data-maturity

Davies, T., & Fernando, P., & Alonso, J. M. (2013). *Researching the emerging impacts of open data: ODDC conceptual framework.* World Wide Web Foundation. http://www.opendataresearch.org/sites/default/files/posts/Researching%20the%20emerging%20impacts%20of%20open%20data.pdf

Deloitte. (2012a). *Open data: Driving growth, ingenuity and innovation.* https://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-data-driving-growth-ingenuity-and-innovation.pdf

Deloitte. (2012b). *Open growth stimulating demand for open data in the UK: A briefing note from Deloitte Analytics.* http://www2.deloitte.com/content/dam/Deloitte/uk/Documents/deloitte-analytics/open-growth.pdf

Fumega, S. (2014). *Opening cities: Open data in Buenos Aires, Montevideo and Sao Paulo: A report for the City of Buenos Aires Open Government Data Initiative.* World Wide Web Foundation. http://opendataresearch.org/sites/default/files/publications/Final-Opening%20Cities%20-%20Buenos%20Aires%20final%20report.pdf

Granickas, K. (2013). *Understanding the impact of releasing and re-using open government data.* European Public Sector Information Platform. http://www.epsiplatform.eu/sites/default/files/2013-08-Open_Data_Impact.pdf

Gruen, N., Houghton, J., & Tooth, R. (2014). *Open for business: How open data can help achieve the G20 growth target.* https://www.omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf

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

Harrison, T., Pardo, T., Cresswell, A., & Cook, M. (2011). *Delivering public value through open government.* The Research Foundation of State University of New York. http://www.ctg.albany.edu/publications/issuebriefs/opengov_pubvalue.pdf

Heimstädt, M., Saunderson, F., Heath, T. (2014). From toddler to teen: growth of an open data ecosystem. *JeDEM-eJournal of eDemocracy and Open Government, 6*(2), 123–135.

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, adoption barriers and myths of open data and open government. *Information Systems Management*, *29*(4), 258–268.

Janssen, M., & Zuiderwijk, A. (2014). Infomediary business models for connecting open data providers and users. *Social Science Computer Review*, *32*(5), 694–711.

Janssen, M., Konopnicki, D., Snowdon, J. L., & Ojo, A. (2017). Driving public sector innovation using big and open linked data (BOLD). *Information Systems Frontiers*, *19*(2), 189–195.

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). Generating value from open government data. *Proceedings of the 34th International Conference on Information Systems (ICIS).* https://aisel.aisnet.org/icis2013/proceedings/GeneralISTopics/5/

Johnson, P. A., & Greene, S. (2017). Who are government Open Data infomediaries? A preliminary scan and classification of open data users and products. *Journal of the Urban & Regional Information Systems Association*, *28*(1), 9–18.

Johnson, P. A., Sieber, R., Scassa, T., Stephens, M., & Robinson, P. (2017). The cost(s) of geospatial open data. *Transactions in GIS*, *21*(3), 434–445.

Johnson, P. A. (2016). Reflecting on the success of open data: How municipal government evaluates their open data programs. *International Journal of E-Planning Research*, *5*(3), 1–13.

Johnson, P. A., (2017). Models of direct editing of government spatial data: Challenges and constraints to the acceptance of contributed data. *Cartography and Geographic Information Science*, *44*(2), 128–38.

Johnson, P. A. (2019). Disintermediating government: The role of open data and smart infrastructure. *Proceedings of the 52nd Hawaii International Conference on System Sciences.* https://doi.org/10.24251/HICSS.2019.346

Kassen, M. (2013). A promising phenomenon of open data: A case study of the Chicago open data project. *Government Information Quarterly*, *30*(4), 508–513. http://www.sciencedirect.com/science/article/pii/S0740624X13000683

Latif, A., Saeed, A. U., Hoefler, P., Stocker, A., & Wagner, C. (2009). *The linked data value chain: A lightweight model for business engineers*. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.181.950&rep=rep1&type=pdf

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E.A. (2013). *Open data: Unlocking innovation and performance with liquid information*. McKinsey & Company. http://www.mckinsey.com/insights/business_technology/open_data_unlocking_innovation_and_performance_with_liquid_information

McHugh, B. (2013). Pioneering open data standards: The GTFS story. In B. Goldstein & L. Dyson (Eds.), *Beyond transparency: Open data and the future of civic innovation*, (pp. 125–135). Code for America Press.

Plu, J., & Scharffe, F. (2012). Publishing and linking transport data on the web: Extended version. *Proceedings of the First International Workshop on Open Data*, (pp. 62–69). ACM Press.

Ruback, L., Casanova, M. A., Raffaetà, A., Renso, C., & Vidal, V. (2016). Enriching mobility data with linked open data. *Proceedings of the 20th International Database Engineering & Applications Symposium*, (pp. 173–182). ACM Press.

Ruppert, E. (2015). Doing the transparent state: Open government data as performance indicators. In R. Rottenburg, S. E. Merry, S.-J. Park, & J. Mugler (Eds.), *The World of Indicators* (pp. 127–150). Cambridge University Press.

Ruvalcaba-Gomez, E. A., Criado, J. I., & Gil-Garcia, J. R. (2018). Discussing open government as a concept: A comparison between the perceptions of public managers and current academic debate. ACM Press.

Sangiambut, S., & Sieber, R. (2017). The civic open data and crowdsourcing app ecosystem: Actors, materials, and interventions. *Journal of the Urban & Regional Information Systems Association*, *28*(1).

Sayogo, D. S., Pardo, T. A., & Cook, M. (2014). A framework for benchmarking open government data efforts. *47th Hawaii International Conference on System Science.* http://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=6758838

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(3), 308–315.

Sunlight Foundation. (2010). *Ten principles for opening up government information.* http://sunlightfoundation.com/policy/documents/ten-open-data-principles/

Ubaldi, B. (2013). Open government data: Towards empirical analysis of open government data initiatives. *OECD Working Papers on Public Governance*, *22*, 1–61. http://dx.doi.org/10.1787/5k46bj4f03s7-en

Vickery, G. (2011). *Review of recent studies on PSI reuse and related market developments*. European Commission. http://ec.europa.eu/newsroom/document.cfm?doc_id=1093

Wang, V., & Shepherd, D. (2020). Exploring the extent of openness of open government data–A critique of open government datasets in the UK. *Government Information Quarterly*, *37*(1), 101405.

Yang, T. M., & Wu, Y. J. (2021). Looking for datasets to open: An exploration of government officials' information behaviors in open data policy implementation. *Government Information Quarterly*, *38*(2), 101574.

Zuiderwijk, A., Janssen, M., Choenni, S., Meijer, R., & Alibaks, R. S. (2012). Socio-technical impediments of open data. *Electronic Journal of e-Government*, *10*(2), 156–172.

Zuiderwijk, A., Helbig, N., Gil-García, J. R., & Janssen, M. (2014). Special issue on innovation through open data: Guest editors' introduction. *Journal of Theoretical and Applied Electronic Commerce Research*, *9*(2), 1–13.

About the Authors

Peter A. Johnson is Associate Professor in the Department of Geography and Environmental Management at the University of Waterloo. His research expertise is in the application and evaluation of geospatial technologies within government, including open data, the geospatial web, mobile devices, crowdsourcing, participatory geomatics, and volunteered geographic information.

Christine Varga is a recent graduate from the Economics program at the University of Waterloo. This work presents the findings of her co-operative work term with the Canadian Open Data Exchange (www.codx.ca).