Received: 10 January 2017 | Revised: 31 March 2017 | Accepted: 2 April 2017

DOI: 10.1111/tgis.12283

RESEARCH ARTICLE

WILEY Transactions in GIS

The Cost(s) of Geospatial Open Data

AQ1 Peter A. Johnson[1] | Renee Sieber[2] | Teresa Scassa[3] | Monica Stephens[4] | Pamela Robinson[5]

[1] Department of Geography and Environmental Management, University of Waterloo

[2] Department of Geography and School of Environment, McGill University

[3] Faculty of Law, University of Ottawa

[4] Department of Geography, University at Buffalo - SUNY

[5] School of Urban and Regional Planning, Ryerson University

Correspondence
Peter A. Johnson, Department of Geography and Environmental Management, University of Waterloo,
Email: peter.johnson@uwaterloo.ca

Abstract

The provision of open data by governments at all levels has rapidly increased over recent years. Given that one of the dominant motivations for the provision of open data is to generate 'value', both economic and civic, there are valid concerns over the costs incurred in this pursuit. Typically, costs of open data are framed as internal to the data providing government. Building on the strong history of GIScience research on data provision via spatial data infrastructures, this paper considers both the direct and indirect costs of open data provision, framing four main areas of indirect costs; citizen participation challenges, uneven provision across geography and user types, subsidy of private sector activities, and the creation of inroads for corporate influence on government. These areas of indirect cost lead to the development of critical questions, including constituency, purpose, enablement, protection, and priorities. These questions are posed as a guide to governments that provide open data in addressing the indirect costs of open data.

KEYWORDS
open data, open government, spatial data infrastructures, technology adoption, technology impacts

1 | INTRODUCTION

GIS experts have borne witness to a period of technological and data-driven change. Traditionally, accessing government data was expensive and complicated, due to proprietary software and data, whereas it is now increasingly ordinary for the public to expect government data sets to be open. This move towards the provision of government data as open, represents a significant change in who can access government data, and under what terms. Open data are government data typically provided for free, in a machine readable format, and with minimal restrictions on reuse (Janssen, Charalabidis, and Zuiderwijk, 2012). Open data is often framed as an untapped resource for tech startup entrepreneurs to drive business models (Manyika, Chui, Groves, Farrell, Van Kuiken, and Doshi, 2013). Governments also may see themselves as open data customers, with open data often posited as advantageous for internal business intelligence (Sieber and Johnson, 2015), and as a way to share information horizontally across levels of government, between jurisdictions, and with the public (Robinson and Johnson, 2016). Most open data contain a fundamental

geospatial component, for example, in the location of crime reports, transit routes, and building permits. Open data form the backbone of base maps, providing the canvas for visualization and analysis of secondary data. Cities also collect a wide variety and ever-increasing volume of real-time, spatially-referenced data on equipment (e.g., the location of public transit or emergency vehicles) and citizen activities (e.g., data from traffic cameras) (Scassa and Diebel, 2016). These spatial data, although often just one component of broader open data catalogues, are the data most often provided by governments and have significant commercial and social value (Janssen et al., 2012).

The current push towards a culture of "openness" in government is often positioned as part of "open government" efforts, with transformative aspirations. In addition to claims of economic benefit, the proposed benefits of open data include greater transparency and accountability, improved data sharing, and increased citizen access to government decision-making (Janssen et al., 2012). Releasing data is touted as an engine of a transformed public engagement, transparency, and accountability (Yu and Robinson, 2012). In cities, open data can be used to construct and to challenge arguments around the nature, character and use of space, with non-profit groups using open data to build arguments for social change (Kassen, 2013). The manifestations of this "openness" and its end results can vary considerably. Recent work highlights this drive for openness and the economic, social, and political tensions created as a result (Bates, 2014; Sieber and Johnson, 2015). Zuiderwijk and Janssen (2014) identify negative effects accompanying the internal government efforts to open data, as well as the management of open data. Their government-side analysis presents a view of open data as technically and organizationally difficult for governments to produce, buffeted by competing institutional goals and regulations. We build on this government-focused assessment of the negative effects of open data by widening our scope to include the implications of open data use more broadly.

Throughout the government, the private sector, and civil society, there is wide-ranging interest in the value of open data, specifically their access, use, and exploitation. As open data communities mature, there comes the recognition that "open" is neither cheap, easy, nor necessarily always positive (Sieber and Johnson, 2015). There is increasing pressure to rationalize government services and reduce costs accompanied by rapid innovations in technologies like web mapping, location-based services, and advanced data analytics. Critical questions arise about the costs incurred through the provision of open data by government. We argue that such costs are both direct (fiscal) and indirect (reflecting unanticipated adverse social or economic impacts). We also note that the lens of "costs" remains only one way to view the emergence and impact of open data, but that this lens is of particular relevance to those governments which provide open data, and to the growing communities of data users. From both provision and use perspectives, there is an often-unexplored nuance to *how* open data are provided, and to *which goals* they contribute.

Due to the spatial dimension of significant amounts of government open data, GIScience provides a unique frame for investigating technological implementation and data sharing, applying data (e.g., geovisualization, spatial analysis), and understanding the consequences of such information (e.g., marginalization of certain populations). This article builds upon investigations of the infrastructures needed to open and share spatial data (Harvey and Tulloch, 2006; Kulk, and Van Loenen, 2012), the early cost/benefit work that evaluated effectiveness and efficiency gains of GIS (Gillespie, 1994; Nedovic-Budic, 1998), and the role of organizational factors in technology adoption (Pinto and Azad, 1994). Through this literature, we focus on identifying the external and often unintended impacts of open data provision (Davies, Janssen, and Schieferdecker, 2014; Sieber and Johnson, 2015). This article presents a synthesis of research findings from four years of collaboration between the authors. This research, funded by [insert organization after peer review is complete], has engaged local government partners across Canada to examine how open data, and in particular geospatial open data, can impact the relationship between government, its citizens and the private sector. As this research has evolved, this theme of the costs of open data has emerged. Our research suggests that "spatial is special" when it comes to open data, including the spatiality of digital divides (Stephens, 2013), and that open data research, which draws upon a different set of assumptions and literatures, would benefit from the lessons learned in GIScience. This article creates a framework that details direct and indirect costs created through the process of government open data provision, drawn from the perspectives of our key informants in municipal governments. Many of the indirect costs are external to the providing organization and are not always clear or straightforward. We use this

discussion of indirect costs to provide four key considerations for government open data provision which encapsulate these indirect costs, as a contribution to the ongoing maturation of the open data movement.

2 | THE DIRECT COSTS OF OPEN GEOSPATIAL DATA

Recent work on the costs of open data typically focus on those borne by governments in the process of providing data (Conradie and Choenni, 2014; Janssen et al., 2012; Zuiderwijk and Janssen, 2014). These direct costs of open data provision include the resource costs to collect the data, to prepare raw data for open publication, to facilitate data sharing via an online portal, and to maintain and update datasets (Johnson, 2016). Costs at each of these steps often manifests through demands on staff time (Chan, Johnson, and Shookner, 2016). For example, data collection is governed by standards and organizational norms that ensure data is appropriate for internal use, though it may remain incomprehensible to external users. This transformation from native format to one that is suitable for "mass" consumption can require several manipulations, including the creation of metadata or even directly modifying variables or datasets. Open data providers also may need to develop tools that assist end users to geovisualize or spatially analyze the data. Rendering the data machine-readable may require a transcribing of data to a common file format or a data standard. There may in fact be only limited standards, structures or support that suggest how a specific dataset should be opened (Janssen et al., 2012). Quality of open data sets can vary widely between different units within the same government, limiting the interoperability of the data (Devillers and Jeansoulin, 2006). The need to update and maintain data sets creates ongoing costs over the life cycle of the data. For example, data may be updated with different frequencies, contain different coverage, or be collected at different scales, with each type of collection creating additional costs. This intra-organizational variation between data sets is augmented when open data end users attempt to merge data from multiple governments, only to find incompatibilities that require expensive remediation

It is critical to note that direct costs of open data must also include considerations of individual privacy and confidentiality. In some cases, data providers will need to inspect and scrub data to protect privacy or to ensure that it is free from third party confidential information. Indeed, concerns over privacy are identified as a barrier to opening up data (Huijboom and van den Broek, 2011). Geospatial attributes complicate privacy protection because the data owner needs to assess the ways in which locational information might identify individuals (Scassa, 2010). For example, in the case of crime data, producing a specific location of a crime incident report might identify the victim of that crime. In some cases, decisions may be made to exclude certain categories of data altogether (e.g., certain types of crime such as domestic violence, where the link to a geographic location makes re-identification risk too high) (Scassa 2016). Thus, the data provider would bear the direct costs of investigating and implementing geographic masking techniques necessary to obscure the exact location yet still provide useful information to the general public (Leitner and Curtis, 2004; Armstrong and Rushton, 1999).

Given these direct costs of self-publishing open data, governments are increasingly seeking third party companies to provide hosting, standardization, and analytics for the government and tools to inspect the data, though these also come with financial and control costs. This places open data providing governments in a difficult spot – generating use and value from open data is much more that simply "handing over" raw data to the public, but rather is augmented by the costs of purchasing or developing an open data platform. These challenges are similar to those found in GIS implementation in local government, where training of staff, developing and agreeing on standards, establishing a valuation for geospatial data, connecting data to various user communities, and receiving feedback on that data proved to be durable challenges (Budic, 1994; Nedovic-Budic, 1998; Pinto and Azad, 1994). It is reasonable to expect that this range of direct costs that may have constrained GIS adoption in municipal government may also have an impact on the adoption of open data. For example, the cost barrier to opening data may lead to the prioritization of those open data sets seen to be most useful to constituencies that might include governments, the private sector and civil society. Some governments have actively solicited input as to what data sets they should open (e.g. Ontario, 2015). Given this cost differential between different datasets, governments may avoid opening data sets that would require a more complicated assessment or anonymization process.

3 | INDIRECT COSTS OF OPEN GEOSPATIAL DATA

The direct costs of providing open data are frequently countered with reference to anticipated benefits. These may include the stimulation of innovation, greater efficiencies within government and enhanced citizen engagement (Janssen et al., 2012). It is increasingly evident that these benefits may not be fully realized in all cases, as anticipated benefits may generate their own drawbacks or challenges. We characterize these challenges as indirect costs. In this section, we describe four main categories of indirect costs: open data as citizen participation smoke and mirrors, uneven geographies of open data provision, open data as a public subsidy of private sector enterprise, and open data as an entry for corporate influence in government. For governments that provide open data, these indirect costs must either be addressed as part of an open data strategy or at the very least acknowledged as real potential outcomes of the open data provision process.

3.1 | Open data as citizen participation smoke and mirrors

Citizen participation is considered a cornerstone of a participatory democracy because civil society serves as an accountability check on government (Pateman, 1970; Barber, 2003). One of the touted benefits of open data is that they can increase citizen participation, and potentially, citizen engagement (Janssen et al., 2012; Davies, Janssen, and Schieferdecker, 2014), but there is little evidence that shows open data are direct drivers of citizen participation. There is potential for the release of open data to support government transparency, for example, by exposing government corruption, by allowing journalists to track government contracts and expenditures or to highlight out of date or poor quality data (Janssen, 2012). Despite this potential, open data may enable a kind of "smoke and mirrors" that obscures a government's actual commitment to citizen participation, transparency and accountability (Robinson, Yu, Zeller, and Felten, 2009).

The challenges of using open data as a platform for citizen participation and engagement can be exacerbated where insufficient government resources are deployed to ensure that open data are properly prepared for release. Where data are poorly structured or incomplete and served by a user-unfriendly interface, the user will experience transaction costs in interacting with data. Specialist data formats (e.g., geoJSON, comma separated values) can serve as a barrier restricting use to those with specific skills, training, and software. Through its focus on data, the open data literature reinforces a perspective that, for example, mapping tools are ubiquitous and easy-to-use. In reality, even the simplest visualization and analysis tools like Google Earth may not be simple enough for a broad section of society to use (Johnson, Corbett, Gore, Robinson, Allen, and Sieber, 2015). These data literacy issues pose a challenge in using open data as a vector for civic participation. Open data advocates believe that, if data are provided, they will be used, with the passive provision of more information-inducing behavior change (Craveiro, Machado, and Machado, 2016). Scholars have noted that placing data in an open data catalogue is no guarantee of its use (Gurstein, 2013; Robinson and Johnson, 2016; Robinson and Ward-Mather, 2017; Sieber and Johnson, 2015). We suggest that governments develop an outreach or training mechanism to educate the public on how to access, use, and interpret open data. For example, some local governments are piloting civic innovation offices through the support of organizations like the Bloomberg Foundation. Public libraries across North America are leaders in this data-literacy space, developing data outreach programs (Robinson and Ward-Mather, 2017). New government-community partnerships through organizations such as Code for America, and Civic Hall (New York City) connect government staff with data and technology experts. These types of interventions can help bridge the gap between data provision and civic outcomes.

There are additional concerns where the provision of open data can lead to outcomes that are contrary to general citizen participation and engagement goals, or even introduce specific harms. For example, open data that fail to address privacy concerns could lead to individuals being targeted in ways that are emotionally, financially, or even physically harmful. This may arise where governments take decisions to make public categories of personal information such as the names and addresses of political campaign donors (Scassa, 2014a). It may also arise where information that is made available as open data is not sufficiently anonymized or de-identified (Conroy and Scassa 2015; Offenhuber, 2015). This creates a challenge in balancing competing values such as privacy and transparency in open data, with

J_ID: Customer A_ID: TGIS12283 Cadmus Art: TGIS12283 Ed. Ref. No.: TGIS-2017-JAN-TGIS-ORA-1539.R1 Date: 8-May-17 Stage:

governments refraining from opening data based on fears that such data may be ineptly or improperly used. For example, gun ownership maps produced by journalists using inaccurate government gun registry data, show how the limitations of the data set may not have been understood or appreciated by those who created the maps (Scassa, 2014b). Regardless of opening data or enabling technologies, government may not wish to subject itself to scrutiny or relinquish its power. Moreover, these efforts at citizen participation and engagement occur in an overly technical ecosystem of only somewhat interoperable components (Brandusescu, Sieber, and Jochems, 2015; Sieber, Robinson, Johnson, & Corbett, 2016). In-person activities like civic hackathons attempt to remedy these barriers, with government and civil society acting as intermediaries between the provision of raw open data and participation and engagement goals (Robinson and Johnson, 2016). Research in participatory GIS demonstrates that technology-led citizen engagement requires it to be carefully framed in a context of who is using, and for what purposes, the hardware, software, and data (Sieber, 2006).

There is a widespread assumption that an informed citizen will be an engaged citizen. Graser and Robinson (2016: 6) argue that open data: "need to be viewed as an input into broader open government efforts in which all actions taken, including the release of data, are in pursuit of government's being more transparent, open, accountable, and accessible". To move beyond the typical approach to providing open data as a one-time dump (Sieber and Johnson, 2015), civic engagement processes need to "wrap" open data, acknowledging that engagement efforts are resource and labor-intensive (Graser and Robinson, 2016; Robinson and Johnson, 2016) and already subject to their own barriers to participation with or without the open-data add on. Opening data offers no magical solution to the durable challenges in participatory democracies. For open data producing governments, these considerations of public participation, engagement, and digital literacy should be central components of an open data program, supporting public use, and mitigating the negative impacts of open data.

3.2 | Uneven Geographies of Open Data Provision

As open data is largely provided as a top-down service by individual levels of government. It is unevenly provided across jurisdictions, creating divisions in access to information (Zuiderwijk, Janssen, & Choenni, 2012). The uneven investment in and quality of open data leaves some municipalities behind socially, politically and economically without the transparency and support for innovation provided by access to open data. Consequently, the availability of specific types of open data sets can advantage one jurisdiction over another by supporting services to one population that are not accessible to another, with this data access translating into both material benefits and social privilege. Social scientists have dubbed this disparity of data availability between jurisdictions as "data poverty" or a "data divide". In addition, even if similar data sets are provided between jurisdictions, the contents of those data sets may differ fundamentally. For example, "if certain groups are routinely excluded from datasets, their problems may be overlooked and their communities held back in spite of progress elsewhere" (Castro, 2014, pg 2). As groups are overlooked for inclusion in data, there are costs to society. These types of digital inequalities, from a government standpoint, means uneven abilities for users to access and harness data.

This uneven provision of data echoes the challenges experienced in the creation of spatial data infrastructures and government-to-government data sharing (Nedovic-Budic and Pinto, 2000; Onsrud and Rushton 1995; Onsrud, 2007). For example, when compiling data from different government departments or agencies, issues of georegistration (e.g., projections, coordinate systems), applying quality standards and schema, and synchronizing periodicity of data updates must be negotiated. Omitola et al. (2010) describe the problem with municipalities storing open data in various formats (PDF, HTML, XLS), and on multiple servers. Differences in licensing (or even a lack of licensing), may lead to problems with the interoperability of data sets and increase costs as governments attempt to share data (Mewhort, 2012; Scassa and Diebel, 2016). These technical and organizational examples of unevenness in open data provision create a patchwork of services that cross jurisdictions, making data access, as well as regional analysis using open data, a challenging task.

Unevenness in open data provision is especially evident when comparing urban and rural municipalities. Rural areas typically represent extremes in capacities for technical skills, but they also differ from urban areas in concerns,

J_ID: Customer A_ID: TGIS12283 Cadmus Art: TGIS12283 Ed. Ref. No.: TGIS-2017-JAN-TGIS-ORA-1539.R1 Date: 8-May-17 Stage:

data structures, and ownership. Rural areas are more likely to collect and maintain datasets on natural resources and emergency management (e.g., evacuations in cases of wildfires). Rural municipalities may or may not be incorporated government entities, so the data is coming from different levels of jurisdiction (a village, a county, a province), each of which may have its own licensing arrangements. Adopting a standard to remedy these problems has a cost as well and complicates intra-regional geospatial data sharing (Harvey and Tulloch, 2006; Nedović-Budić and Pinto, 2000).

Unevenness in open data provision is also formalized in the platforms chosen by municipalities to deliver open data. Many cities are shifting from developing their own portals to third party solutions (Center for City Solutions and Applied Research, 2014). Although there are open source alternatives, popular solutions are often proprietary. Stephens (2017) describes Socrata, the popular cloud-based software service used by many municipalities as problematic as it charges governments on a pro-rata basis. Thus, for example, cities with large, poor populations pay more than smaller, wealthier cities. A focus by open data providing governments on financing the mechanics of delivery can come at the expense of nurturing a cooperative culture around data and tools that can lead to greater and more diverse use of open data (Bolukbasi et al., 2013).

3.3 | Open data as a public subsidy of private sector enterprise

Historic discussions about creating an infrastructure for distributing spatial data serve as cautionary tales, as it was expensive to prepare this data for public release but also because the main entities using spatial data were the private sector (Onsrud, 2007). Given the benefits for the private sector in using open data instead of generating or purchasing similar data from other sources (Janssen et al., 2012; Johnson, 2016), this raises questions as to what degree the public sector is subsidizing private sector business models by opening data. If open data is framed as "the new gold", (European Commission, 2011) then it may become difficult for government to justify bearing the cost of continued production as well as expansion and distribution.

Early arguments over the provision of geospatial data by government were grounded in a debate over whether that data should be considered open access or should be sold (Onsrud, 1992a,b). Supporting the view that government geospatial data should be openly accessed as a public good were arguments that government data was already funded by the taxpayers, that open data should be considered a right in a democratic society, and that transparency values dictated release (Onsrud, 1992a). These arguments are perhaps stronger where the data being released as open data is relatively close in format and content to that which is already compiled and used by governments. These arguments are less compelling where data must be compiled or processed in new ways or reformatted, or must be screened for privacy or confidential information prior to being made open. Onsrud (1992b) noted that the high cost of collecting and maintaining government data, combined with the potential for release of data that compromised privacy or security, might support arguments for government maintaining control over its data. Such arguments also support the government's right to distribute data on a cost-recovery basis. The current characterization of open data as a public good that should be freely available, absent any restrictions, revisits these same arguments (Boulton, Rawlins, Vallance, and Walport, 2011).

Where the private sector comes to rely upon open data, it will inevitably push for the release of data sets that most serve its needs. The potential therefore exists for open data to become corporatized, with government resources directed toward "high value" and away from "low value" data sets, where value is understood in commercial terms. High value data sets are those which can form the basis for commercial products or services (such as base geographical data, or real-time transit data), or those which can be combined with other data to produce commercial products or services (such as postal codes or geodemographic data). Open data is increasingly relied upon in big data analytics as well (Kitchin, 2014). Such analytics may absorb a broad range of open data in combination with data from other sources. So-called low value data sets have a less obvious commercial potential. Some of these, such as data sets relating to municipal services or installations, may have greater importance to civil society.

Given the costs of data production and maintenance, a two-tier pricing structure with paid commercial use and free non-commercial use of government open data could provide cost recovery. Further, the rhetoric of open data, which often refers to it as data for which taxpayers have already paid (Coglianese, 2009; Obama, 2009), may obscure

the real costs to government of making data open, masking the true value of the subsidy to the private sector. This tension between government expense and private sector use raises many questions. Should open data be treated as a specialist product and offered on a cost recovery basis? Increasingly, under the smart cities model, access to infrastructure and service data is becoming a revenue stream for cash-strapped governments (Ching and Ferriera, 2015). Given this new-found funding source, it is important that government balance the costs of open data provision with the broad benefits generated, particularly from a civil society perspective. If open data is provided largely for private sector consumption, there must be more than trickle-down benefits for citizens. Barring this, government may move towards recouping at least some of the costs from those high volume data users that turn a profit on the back of open data.

3.4 | Open data as an entry for corporate influence in government

Open data are often presented as a service to the private sector, with the goal of supporting economic development and an innovation agenda (Davies et al., 2014). Providing open data in this manner aligns with the "government as a platform" model proposed by O'Reilly (2011). With this model, governments provide open data as a type of infrastructure, yet largely withdraw from using this open data to provide citizen-facing services, instead leaving the application of data to market forces, following a neoliberal model (Bates, 2014; Longo, 2011; Sieber and Johnson, 2015). In this way, services provided by the private sector using open data become firmly entrenched. Leszczynski (2012) points out that public-private collaborations to add value to geospatial data may ultimately tie governments to the licensing of proprietary software and services. For example, contracts with private sector companies to provide the technological and algorithmic infrastructure to collect and process data may either prevent such data from being shared more broadly (Scassa, 2014a) or may tie the terms of data use to the requirements of the private sector supplier (Scassa and Diebel, 2016). This direct corporate influence in data provision extends to data formats as well. Google's participation in open data standards like the General Transit Feed Specification (GTFS) for transit data (such as route and timetable information) may illustrate the state's "rollback" of government curation, dissemination, analysis, and visualization of spatial data. This creates a greater reliance on private partners to set and promote standards, sharing control over data publication formats, while also acknowledging the benefits in interoperability that come from standardization (Leszczynski, 2012).

Open data also may be used strategically by the private sector in government procurement processes. Bates (2014) demonstrates that some open data can be used by private sector enterprises to gain commercial advantages in developing bids to supply public sector services. In this way, she notes that open data may support privatization agendas, and also may lead to "cherry-picking" regarding those services for which bids should be made and those less profitable ones which should be left to governments to fund. Bates also questions the use of open government data by governments in deliberate and strategic ways to reinforce the neoliberal state. Her research points to the use of open data in "the marketisation of public services and potential privatisation of public assets, the leveraging of financial markets and pharmaceutical industries through selective release of particular datasets, and the embedding of OGD into a broader agenda aimed at (re)building trust in political elites" (Bates, 2014, 394). In a similar vein, a growing number of private sector companies offer services to the government or the private sector that re-package, process, or analyze government data. Whereas to some extent this reflects an innovation agenda that has driven many open data programs; it also raises the spectre of the outsourcing of information production and analysis (Bates, 2014). A critical examination is needed of the role the open data portal companies play as intermediaries shaping how government open data is accessed and used. For example, in the case of visualizations of urban crime data, Scassa (2016) demonstrates how private sector design and implementation of crime mapping platforms impacts the representation of both crime data and the respective roles of citizens and police.

An indirect consequence of this model is that, rather than an open market for data access, the private sector is largely given free reign to create the access points for citizens (e.g., through application development). There are exceptions where government and non-profit agencies support civic services through the use of open data (Sangiambut and Sieber, under review). The current government paradigm of increasing the contracting out of program and service provision has created a system where benefits from the provision of open data as a public good may eventually trickle

down to the citizen, but instead are more commonly embedded in a private sector business model. This paradigm layers on a number of costs to the public. First, taxes paid by citizens to fund the provision of the government data (a direct cost) are augmented by the indirect costs of not having reliable or timely assurances of access to the data they effectively paid for. Simultaneously, the private sector profits (an indirect cost) and then offers a now-private service back to the public at a fee (another direct cost).

4 | CONCLUSIONS: CREATING VALUE, MITIGATING COSTS FROM OPEN DATA PROVISION

The early days of the open data movement challenged the commitment of governments to be "open by default" with their data in machine readable form, free of licensing and other restrictions. As many governments open their data sets, common approaches were shared, helping to resolve technical, policy, and licensing issues (Sieber and Johnson, 2015). Despite promising beginnings and the growing maturity of the open data movement, durable challenges remain and new challenges emerge. The current push towards identifying the types of value generated from open data, whether that value is economic (Charalabidis, Loukis, and Alexopoulos, 2014), social (Jetzek, Avital, and Bjørn-Andersen, 2013), or political (Kassen, 2013), must also acknowledge costs. These costs, both direct and indirect, are generated from the process of open data provision, through the ways that open data is applied and valued, and over time throughout the data life cycle (Sieber and Johnson, 2015). For example, although there is potential for open data to support citizen engagement, these processes are already fraught with challenges that the provision of open data does not clearly resolve. Recent case study research shows challenges with open data platforms, highlighting the exclusionary nature of many municipal open data programs, and how unevenly open data is provided, particularly when comparing urban vs. rural communities (Chan et al., 2016). As a greater variety of datasets are released, these datasets (particularly high-value health and location information) raise critical privacy questions (Scassa, 2010). Lastly, given the highly specialist nature of much open data, the reinforcing of traditional digital divides raises questions of how open data exists as a subsidy to various private sector business models (Longo, 2011; Bates, 2014). While it might be easy, at first read, to frame this critique of costs as overtly negative, instead the authors intend this fresh and honest look at open costs to inform future action that is thoughtful, strategic and pragmatic in response to many of the real challenges that lie ahead.

By identifying the indirect costs of open data, we distil critical questions that can guide the long-term future of open data research that should be acknowledged by governments, driving the future progress of open data provision as it matures beyond a simple download from a website. In creating a positive future for open data provision, government must address key questions of constituency, purpose, enablement, protection, and priorities. For both governments and researchers considering how value can be generated from open data, we propose that questions of costs – both direct and indirect – must accompany those of value in equal measure.

When opening data, government data providers must consider for which *constituency* this open data is intended. As detailed above, durable challenges to accessing data can have little to do with its openness, and are more focused on concerns of format, technical knowledge required, standardization across jurisdictions, as well as data completeness and quality concerns. In addition, the very nature of open data provision lacks tracking mechanisms that would allow government to have a comprehensive understanding of its user base. Simply put, providing data for public use must account for varying definitions of what a "public" is, and not be simply tailored to one constituency, to the exclusion of others.

The *purpose* to which a user community puts open data can be impacted by the way open data is structured and provided and through which specific data sets are made available. For example, the use of open data for business development purposes will demand certain standards, completeness, and frequency of updating. Open data used by civil society for transparency purposes may require a level of personal identification (e.g., city councillor names). In providing open data that is adapted for these uses, government must consider if they are providing the kind and quality of data suitable to meet these purposes.

Open data is not released into a vacuum, but rather to a user base with diverse abilities and goals. When considering the process of *enablement*, government needs to acknowledge that there is a gap between open data provided "as-is" and the ability of a given user community to exploit that data in the creation of value. These enablement costs include additional measures, policies or practices required for the realization of the public purposes listed above (e.g., improved literacy, public outreach, hackathons, and public sector innovation). For example, the enablement needs of the private sector may include seamless, machine readable data in a well-known standard, and with no limits on commercial reuse. Realizing the goal of generating value from open data requires that this gap between data and end use be bridged.

The direct costs of creating, hosting, and sharing open data do not necessarily need always to fall upon the government, but rather there are roles for the private and not-for-profit sectors in supporting open data release. For example, Google created the General Transit Feed Specification (GTFS) (https://developers.google.com/transit/gtfs) and promoted this as a standard for how government could release real-time public transit schedules. This development of a common standard across jurisdictions eased development costs for a variety of end-user applications. Essentially, the government maintained control over the data and the private sector created an open format specification that eases the ability of developers to scale their transit applications across jurisdictions. This type of interaction shows how feedback and support from data user to provider can enable the use of open data and the resulting creation of value.

Government, as data custodians, must consider the *protection* of data sources, whether vital infrastructure or individual citizens. It is the government's role to ensure that methods are in place for the anonymization or aggregation of data to protect privacy, as well as conducting necessary privacy, security, and risk assessments (Cavoukian et al. 2010). For example, what measures or precautions are necessary to protect against the harms of misuse of open data? Given the increasing pressure on the release of location, individual, and health-related data, protection of sensitive data remains a critical concern in the search for open data value.

Lastly, these questions include asking who sets the *priorities* for open data release and update? As the providers of open data, governments must aim to balance the societal benefits of open data with open data's inclusion in private sector business models, while considering the costs of open data, for whom is open data made available, and to what ends is it directed. Prioritizing release of data sets may be complicated, for example, as open data may increase access for technical elites and simultaneously marginalize traditional civic actors (Bates, 2014).

The provision of geospatial open data is continuing to expand. In parallel, researchers are beginning to address many of the nuances of this new form of data provision. GIScience, with its recent history of SDI research and focus on technical, organization, and social factors to spatial data provision and GIS adoption can serve as a resource for open data researchers to better understand how both value and costs are generated. Measuring, evaluating and balancing these costs to open data are fraught with durable challenges, for example balancing open data as a meaningful contributor to citizen engagement with protecting citizen privacy through de-identification of personal data. This article pushes the discussion of open data value beyond broad statements about its value traditionally created through transparency, innovation, or civic activity. Rather, we frame the direct and indirect costs that are generated through both provision and application of open data. Establishing the value of open data for diverse stakeholders requires a move beyond considering just the direct costs to government in providing data, but also to include those indirect costs that impact a broad variety of institutions and sectors of society.

ACKNOWLEDGEMENTS

The authors would like to thank the Social Sciences and Humanities Research Council of Canada (SSHRC) for funding this research via the Geothink.ca Partnership Grant.

REFERENCES

Armstrong, M.P., Rushton, G. and Zimmerman, D. L., (1999). Geographically masking health data to preserve confidentiality. *Statistics in Medicine*, *18*(5):497–525.

Barber, B. (2003). *Strong democracy: Participatory politics for a new age* (2nd Ed.). Berkeley, CA: University of California Press.

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.

Bolukbasi, B., Berente, N., Cutcher-Gershenfeld, J., Dechurch, L., Flint, C., Haberman, M., ... & McElroy, C. (2013). Open data: crediting a culture of cooperation. *Science*, *342*(6162), 1041–4042.

Boulton, G., Rawlins, M.,Vallance, P., & Walport, M. (2011). Science as a public enterprise: The case for open data. *The Lancet*, *377*(9778), 1633–1635.

Brandusescu, A., Sieber, R. E., & Jochems, S. (2015). Confronting the hype: The use of crisis mapping for community development. *Convergence: The International Journal of Research into New Media Technologies*, *22*(6), 616–632.

Budic, Z. D. (1994). Effectiveness of geographic information systems in local planning. *Journal of the American Planning Association*, *60*(2), 244–263.

Castro, D. (2014). *The rise of data poverty in America*. Washington, DC: Center for Data Innovation (retrieved from http://www2.datainnovation.org/2014-data-poverty.pdf).

Cavoukian, A., Taylor, S., & Abrams, M. E. (2010). Privacy by design: Essential for organizational accountability and strong business practices. *Identity in the Information Society*, *3*(2), 405–413.

Center for City Solutions and Applied Research (2014). *City open data policies*. Washington, DC: National League of Cities (retrieved from http://www.nlc.org/Documents/Find%20City%20Solutions/City-Solutions-and-Applied-Research/CSAR%20Open%20Data%20Report%20FINAL.pdf).

Chan, M., Johnson, P., & Shookner, M. (2016). Assessing the use of government open data and the role of data infomediaries: The case of Nova Scotia's Community Counts Program. *eJournal of eDemocracy and Open Government*, *8*(1), 1–27.

Charalabidis, Y., Loukis, E., & Alexopoulos, C. (2014). Evaluating second generation open government data infrastructures using value models (pp. 2114–2126). In *Proceedings of the Annual Hawaii International Conference on Systems Sciences*. Waikoloa, HI.

Ching, T-Y., & Ferreira J. (2015). Planning Support Systems and Smart Cities. In S. Geertman, J. Ferreira, Jr., R. Goodspeed, & J. Stillwell (Eds.), *Smart cities: Concepts, perceptions and lessons for planners* (pp. 145–168). Berlin, Germany: Springer Lecture Notes in Geoinformation and Cartography.

Coglianese, C. (2009). The transparency president? The Obama administration and open government. *Governance*, *22*(4), 529–544.

Conradie, P., & Choenni, S. (2014). On the barriers for local government releasing open data. *Government Information Quarterly*, *31*, S10–S17.

Conroy, A. and Scassa, T. (2015). Promoting transparency while protecting privacy in open government in Canada. *Alberta Law Review*, *53*(1), 175–206.

Craveiro, G. S., Machado, J. A., & Machado, J. S. (2016). The use of open government data to citizen empowerment. In *Proceedings of the 9th International Conference on Theory and Practice of Electronic Governance* (pp. 398-399). Montevideo, Uruguay: ACM.

Davies, T., Janssen, M., & Schieferdecker, I. (2014). Open data: Growing up and getting specific. *eJournal of eDemocracy & Open Government*, *6*(1), i–iii.

Devillers, R., & Jeansoulin, R. (Eds.). 2006. *Fundamentals of spatial data quality*. New York, NY: Wiley-ISTE.

European Commission (2011). Data is the new gold. Opening remarks, Press Conference on Open Data Strategy. Brussels. Retrieved from https://ec.europa.eu/digital-single-market/en/news/data-new-gold.

Gillespie, S. R. (1994). Measuring the benefits of GIS use: Two transportation case studies. *URISA Journal*, *6*(2), 62–67.

Graser, D., & Robinson P. (2016). *A recipe for fiscal trust*. Toronto, ON: University of Toronto Institute for Municipal Finance and Governance Perspectives Paper No. 13 (retrieved http://munkschool.utoronto.ca/imfg/uploads/350/imfgperspectives_no13_fiscaltrust_graser_robinson_july_21_2016.pdf).

Gurstein, M. (2013). *Should "Open Government Data" be a product or a service (and why does it matter)?* Gurstein's Community Informatics (3 February) (retrieved from https://gurstein.wordpress.com/2013/02/03/is-open-government-data-a-product-or-a-service-and-why-does-it-matter/).

Harvey, F., & Tulloch, D. (2006). Local-government data sharing: Evaluating the foundations of spatial data infrastructures. *International Journal of Geographical Information Science*, *20*(7), 743–768.

Huijboom, N., & Van den Broek, T. (2011). Open data: An international comparison of strategies. *European journal of ePractice*, *12*(1), 4–16.

Janssen, K. (2012). Open government data and the right to information: Opportunities and obstacles. *Journal of Community Informatics*, *8*(2).

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.

Jetzek, T., Avital, M., & Bjørn-Andersen, N. (2013). Generating value from open government data. In *Proceedings of the 34th International Conference on Information Systems*. Milan, Italy: ICIS.

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), 1–12.

Johnson, P. A., Corbett, J. M., Gore, C., Robinson, P., Allen, P., & Sieber, R. (2015). A web of expectations: Evolving relationships in community participatory Geoweb projects. *ACME: An International E-Journal for Critical Geographies*, *14*(3), 827–848.

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.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society*, *1*(1).

Kulk, S., & Van Loenen, B. (2012). Brave new open data world. *International Journal of Spatial Data Infrastructures Research*, *7*.

Leitner, M. & Curtis, A. (2004). Cartographic guidelines for geographically masking the locations of confidential point data. *Cartographic Perspectives*, *49*, 22–39.

Leszczynski, A. 2012. Situating the geoweb in political economy. *Progress in Human Geography*, *36*(1), 72–89.

Longo, J. (2011). # OpenData: Digital-era governance thoroughbred or new public management Trojan horse?. *Public Policy & Governance Review*, *2*(2), 38.

Manyika, J., Chui, M., Groves, P., Farrell, D., Van Kuiken, S., & Doshi, E. A. (2013). *Open data: Unlocking innovation and performance with liquid information*. Los Angeles, CA: McKinsey Global Institute.

Mewhort, K. (2012). *Creative commons licences: Options for Canadian open data providers*. Ottawa, ON: Faculty of Law, University of Ottawa.

Nedović-Budić, Z. (1998). The impact of GIS technology. *Environment & Planning B*, *25*(5), 681–692.

Nedović-Budić, Z. & Pinto, J. (2000). Information sharing in an inter-organizational GIS environment. *Environment & Planning B*, *27*, 455–474.

Obama, B. (2009). *Transparency and open government*. Washington, DC: White House Memorandum for the Heads of Executive Departments & Agencies.

Offenhuber, D. (2015). Infrastructure legibility: A comparative analysis of open311-based citizen feedback systems. *Cambridge Journal of Regions, Economy & Society*, *8*(1), 93–112.

Omitola T, CL Koumenides, IO Popov, Y Yang, M Salvadores, M Szomszor, . . ., & Shadbolti, N. (2010). Put in your postcode, out comes the data: A case study. In L. Aroyo, G. Antoniou, E. Hyvönen, A. ten Teije, H. Stuckenschmidt, L. Cabral, & T. Tudorache (Eds.), *Semantic Web: Research and Applications, The Semantic Web: Research and Applications: 7th Extended Semantic Web Conference, ESWC 2010, Heraklion, Crete, Greece, May 30 - June 2, 2010, Proceedings, Part I* (pp. 318-332). Berlin, Germany: Springer.

Onsrud, H. J. (2007). *Research and theory in advancing spatial data infrastructure concepts*. Redlands, CA: Esri Press.

Onsrud, H. J. (1992a) In support of open access for publicly held geographic information. *GIS Law*, *1*(1), 3–6.

Onsrud, H. J. (1992b) In support of cost recovery for publicly held geographic information. *GIS Law*, *1*(2): 1–6.

Onsrud, H. J., & Rushton, G. (Eds.) (1995). *Sharing geographic information*. New Brunswick, NJ: Center for Urban Policy Research.

Ontario. 2015. Sharing government data (Published October 2, 2015, updated March3, 2017). Retrieved from https://www.ontario.ca/page/sharing-government-data.

O'Reilly, T. (2011). Government as a Platform. *innovations*, *6*(1), 13–40.

Pateman, C. (1970). *Participation and democratic theory*. Cambridge, UK: Cambridge University Press.

Pinto, J. K., & Azad, B. (1994). The role of organizational politics in GIS implementation. *Journal of the Urban and Regional Information Systems Association*, *6*(2), 1.

Robinson, D. G., Yu, H., Zeller, W. P., & Felten, E. W. (2009). Government data and the invisible hand. *Yale Journal of Law & Technology*, *11*, 160.

Robinson, P. J., & Johnson, P. A. (2016). Civic hackathons: New terrain for local government-citizen interaction? *Urban Planning*, *1*(2), 65–74.

Robinson, P. J., & Ward-Mather, L. (2017). Open data community maturity: Libraries as civic infomediaries. *URISA Journal* (submitted).

Sangiambut, S., & Sieber, R. E. (2017). The open data app ecosystem: Actors, materials, and interventions. *URISA Journal* (submitted).

Scassa, T. (2010). Geographical information as personal information. *Oxford University Commonwealth Law Journal*, *10*(2), 185–214.

Scassa, T. (2014a). Public transit data through an intellectual property lens: Lessons about open data. *Fordham Urban Law Journal*, *41*, 1759–1810.

Scassa, T. (2014b). Privacy and open government. *Future Internet*, *6*(2), 397–413.

Scassa, T. (2016). Police Service crime mapping as civic technology: A critical assessment. *International Journal of E-Planning Research*, *5*(3), 13–27.

Scassa, T., & Diebel, A. (2016). Open or closed? Open licensing of real-time public sector transit data. *Journal of e-Democracy*, *8*(2), 1–20.

Sieber, R. E. (2006). Public participation geographic information systems: A literature review and framework. *Annals of the American Association of Geographers*, *96*(3): 491–507.

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.

Sieber. R. E., Robinson. P. J., Johnson P. A., & Corbett. J. (2016) Doing public participation on the Geospatial Web. *Annals of the American Association of Geographers*, *106*(5), 1030–1046.

Stephens, M. (2013). Gender and the GeoWeb: Divisions in the production of user-generated cartographic information. *GeoJournal*, *78*(6), 981–996.

Stephens, M. (2017). Inequalities across scales of OpenData. *Presentation at the Annual Meeting of the Association of American Geographers*. Boston, MA.

Yu, H., & Robinson, D. (2012). *The new ambiguity of 'Open Government'*. Princeton, NJ: Princeton CITP/Yale ISP Working Paper (retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2012489).

Zuiderwijk, A., Janssen, M., & Choenni, S. (2012). Open data policies: Impediments and challenges. In *Proceedings of the 12th European Conference on E-government* (pp. 794-800). Barcelona, Spain.

Zuiderwijk, A., & Janssen, M. (2014). The negative effects of open government data: Investigating the dark side of open data (pp. 147–152). In Proceedings of the 15th Annual International Conference on Digital Government Research (pp. 147-152). Aguascalientes, Mexico: ACM.

How to cite this article: Johnson PA, Sieber R, Scassa T, Stephens M, Robinson P. The Cost(s) of Geospatial Open Data. *Transactions in GIS*. 2017;00:1–12. https://doi.org/10.1111/tgis.12283

AUTHOR QUERY FORM

Dear Author,
During the preparation of your manuscript for publication, the questions listed below have arisen. Please attend to these matters and return this form with your proof.
Many thanks for your assistance.

Query References	Query	Remarks
AQ1	Please confirm that given names (red) and surnames/family names (green) have been identified correctly.	

Please confirm that the funding sponsor list below was correctly extracted from your article: that it includes all funders and that the text has been matched to the correct FundRef Registry organization names. If a name was not found in the FundRef registry, it may be not the canonical name form or it may be a program name rather than an organization name or it may be an organization not yet included in FundRef Registry. If you know of another name form or a parent organization name for a not found item on this list below, please share that information.

FundRef name	FundRef Organization Name (Country)	FundRef DOI	Grant IDs
Social Sciences and Humanities Research Council of Canada	Social Sciences and Humanities Research Council of Canada	10.13039/501100000155	Geothink.ca