Mediating Open Data: Providers, Portals, and Platforms

Who Are Government Open Data Infomediaries? A Preliminary Scan and Classification of Open Data Users and Products

Exploring Open Data Perspectives from Government Providers in Western Canada

Open Data Community Maturity: Libraries as Civic Infomediaries

The Geospatial Contents of Municipal and Regional Open Data Catalogs in Canada

The Civic Open Data and Crowdsourcing App Ecosystem: Actors, Materials, and Interventions
Check out the projects section on the GISCorps website (www.giscorps.org) for a comprehensive look at past projects, current projects, and future project needs.

"GISCorps volunteers were instrumental in establishing the Australian Red Cross with a spatial mapping capability to be used during emergencies. The value the GISCorps volunteers brought to the project was proven in the success of the mapping products developed during the 2011 QLD Floods and cyclones."

- Emergency Services Coordinator, Australian Red Cross
CONTENTS

7 Mediating Open Data: Providers, Portals, and Platforms
   Edited by Victoria Fast and Claus Rinner

9 Who Are Government OpenData Infomediaries? A Preliminary Scan and Classification of Open Data Users and Products
   Peter A. Johnson and Sarah Greene

19 Exploring Open Data Perspectives from Government Providers in Western Canada
   Mark Gill, Jon Corbett, and Renée Sieber

31 Open Data Community Maturity: Libraries as Civic Infomediaries
   Pamela Robinson and Lisa Ward Mather

39 The Geospatial Contents of Municipal and Regional Open Data Catalogs in Canada
   Edgar Baculi, Victoria Fast, and Claus Rinner

49 The Civic Open Data and Crowdsourcing App Ecosystem: Actors, Materials, and Interventions
   Suthee Sangjambut and Renée Sieber
EDITORIAL BOARD

URISA Journal Editor
Editor-in-Chief
Jochen Albrecht, PhD, Associate Professor for Computational and Theoretical Geography, Hunter College, New York, NY

Thematic Editors:

Sustainability Analysis and Decision Support Systems:
Timothy Nyerges, University of Washington

Participatory GIS and Related Applications:
Laxmi Ramasubramanian, Hunter College

Social, Economic, Governance and Political Sciences:
Francis Harvey, University of Minnesota

GIScience:
Harvey Miller, University of Utah

Urban and Regional Systems and Modeling:
Itzhak Benenson, Tel Aviv University

Editorial Board

Peggy Agouris, Center for Earth Observing and Space Research, George Mason University, Virginia

David Arctur, Open Geospatial Consortium

Michael Batty, Centre for Advanced Spatial Analysis, University College London (United Kingdom)

Kate Beard, Department of Spatial Information Science and Engineering, University of Maine

Yvan Bédard, Centre for Research in Geomatics, Laval University (Canada)

Itzhak Benenson, Department of Geography, Tel Aviv University (Israel)

Al Butler, GISP, Milepost Zero

Barbara P. Buttenfield, Department of Geography, University of Colorado

Keith C. Clarke, Department of Geography, University of California-Santa Barbara

David Coleman, Department of Geodesy and Geomatics Engineering, University of New Brunswick (Canada)

Paul Cote, Graduate School of Design, Harvard University

David J. Cowen, Department of Geography, University of South Carolina

William J. Craig, GISP, Center for Urban and Regional Affairs, University of Minnesota

Robert G. Cromley, Department of Geography, University of Connecticut

Michael Gould, Environmental Systems Research Institute

Klaus Greve, Department of Geography, University of Bonn (Germany)

Daniel A. Griffith, Geographic Information Sciences, University of Texas at Dallas

Francis J. Harvey, Department of Geography, University of Minnesota

Bin Jiang, University of Gävle, Sweden

Richard Klosterman, Department of Geography and Planning, University of Akron

Jeremy Mennis, Department of Geography and Urban Studies, Temple University

Nancy von Meyer, GISP, Fairview Industries

Harvey J. Miller, Department of Geography, University of Utah

Zorica Nedovic-Budic, School of Geography, Planning and Environmental Policy, University College, Dublin (Ireland)

Timothy Nyerges, University of Washington, Department of Geography, Seattle

Harlan Onsrud, Spatial Information Science and Engineering, University of Maine

Zhong-Ren Peng, Department of Urban and Regional Planning, University of Florida

Laxmi Ramasubramanian, Hunter College, Department of Urban Affairs and Planning, New York City

Carl Reed, Open Geospatial Consortium

Claus Rinner, Department of Geography, Ryerson University (Canada)

Monika Sester, Institute of Cartography and Geoinformatics, Leibniz Universität Hannover, Germany

David Tulloch, Department of Landscape Architecture, Rutgers University

Stephen J. Ventura, Department of Environmental Studies and Soil Science, University of Wisconsin-Madison

Barry Wellar, Department of Geography, University of Ottawa (Canada)

Lyna Wiggins, Department of Planning, Rutgers University

F. Benjamin Zhan, Department of Geography, Texas State University-San Marcos
THANK YOU

The first issue of the URISA Journal was published in 1989, and printed by the University of Wisconsin Press. The Editors were Kenneth Dueker, D. David Moyer, and Bernard Niemann, Jr.

That first issue was about 140 pages in length and included refereed and Horwood Critique Prize papers, feature articles, a featured map, book and software reviews:

The URISA Community owes a huge debt of gratitude to the editors, guest editors, members of the editorial review board, and authors for their amazing commitment to the publication over the years.

Editor 1999-2003: Harlan Onsrud
Editor 2003-2006: Steven Ventura
Editor 2006-2011: Jochen Albrecht
Editor 2011-2015: Piyushimita (Vonu) Thakuriah
Editor 2015-2017: Jochen Albrecht

Take some time to look through the archives and see just how far we have come: http://www.urisa.org/resources/urisa-journal/
Final Letter from the Editor

As you read these lines, I regret to inform you that this will be the last volume and issue of the URISA Journal. For many of us, it feels as if the Journal has been around forever (or at least as long as URISA itself). In part, this is because URISA was quite academic in character when the organization was founded—the profession, and certainly the notion of a GIS professional, did not exist in the early days. Since URISA’s inception, numerous academics, including Bob Aangeenbrug, Will Craig, Ken Duecker, Britt Harris, Edgar Horwood, Bill Huxhold, Joe Ferreira, Barry Wellar, and Lyna Wiggins, served as presidents of our organization, providing the leadership and momentum to create the profession that we now are. Journal editors such as Ben Nieman, Harlan Onsrud, Steve Ventura, board members such as Nancy Obermeyer, and successive generations of the Journal’s editorial board provided invaluable service in shaping the quality and reach of the Journal.

The academic intellectual roots that resulted in the creation of our Journal along with its peer-review process provided an academic seal of approval to authors, readers, and URISA members and celebrated high-quality GIS research and practice. URISA originally was modeled after the Regional Science Organization; URISA conferences and later the papers published in the early years of the Journal were products of serious and rigorous scholarship—often practice-oriented, no doubt, but really contributing to what in later years was to be dubbed geographic information science. Even now, the majority of our mission statement’s seven bullet points have the word science in them, and two other points use science in the explanatory body of their bullet header.

With the development and growth of the GIS profession, the interests of our readers have moved away from the original focus on foundational concepts of GIScience and towards the plethora of GIS application areas that spans every aspect of humanity. Simultaneously, the academic GIS scene has developed its own range of GIS journals that include many subspecializations. The offerings available to GIS professionals outside of academia also have grown during this time. The dramatic growth of journal offerings as outlets for academic and scholarly work limits the reach of the URISA Journal to academic scholars while simultaneously making it less interesting to GIS professionals who are looking for reports about best practices and case studies associated with the complexities of GIS use, implementation, or policy.

The academic representation in URISA has dropped significantly, now less than 5 percent of the overall membership. This decline has had serious consequences for the viability of the Journal. For a while, the Journal kept its academic profile, and many professional members were proud to belong to an organization that published such highly ranked articles. The next trend was the influx of international submissions. The URISA Journal was and is uniquely positioned to bridge the increasing gap between theory and practice and our friends from AURISA (now SSSI) and a whole cohort of former regional scientists now geospatial infrastructure architects from The Netherlands helped to justify our association's name URISA International. These articles, plus the general themes of building cadasters and how to engage the public (PPGIS), were the hallmark of our Journal in the 1990s and naughts, propelling our journal in 2008 to number 14 out of 84 GIScience journals evaluated in a Transactions in GIS article (Caron et al. 2008). The next generation of authors reported how these lessons were applied in the developing world, especially the Caribbean, the Middle East, and India.

In 2006, when I took over as general editor, I appealed to the largely professional and U.S.–based membership to submit articles that discussed best practices and lessons learned (and to be shared) that were critical and reflective, in other words, narratives that were qualitatively different from those in trade magazines. In the years since, submitting authors have shown difficulty heeding this request. As I write these lines, one corrective nudge immediately raises its head. Although our membership is mostly U.S.–based, our Canadian brethren contributed more than their fair share to GIS in general, academic contributions, and even articles published in our Journal. It is therefore no coincidence that this last issue is a Canadian one. I am thanking Victoria Fast and Claus Rinner for corralling their colleagues and managing a prolonged review process.

As before, I encourage each of you to send me your feedback, comments, etc. You can reach me at jochen@hunter.cuny.edu.

Cheers and tears,
Jochen

Mediating Open Data: Providers, Portals, and Platforms

Edited by Victoria Fast and Claus Rinner

There has been incredible growth in the presence and substance of open data initiatives worldwide. Governments at all levels of jurisdiction face increasing pressure to release public-sector data in machine-readable formats with minimal restrictions. However, the release of data, particularly government data, is not a trivial task, and requires significant mediation to be realized. The forces that create, shape, and transform the open data landscape can be defined as infomediaries—intermediaries that support the creation and sharing of digital information. Infomediaries, both human and nonhuman, negotiate the gap between open data providers and end-users, and can take the form of service providers, portals, and platforms.

The purpose of this special issue is to explore a range of infomediary roles, including government (articles by Johnson and Greene; Gill, Corbett and Sieber), libraries (Robinson and Mather), data formats (Baculi, Fast, and Rinner), and software (Sangiambut and Sieber). The cases in this special issue focus on open data mediation within Canada, one of the leading nations of the open data movement and a key contributor to geographic information systems (GIS) and geospatial data infrastructure development. The legacy of decades of work in building better systems for data sharing in Canada has led to one of the world’s most pervasive, multiscale open data strategies, informing urban and regional information systems nationally and internationally.

The special issue begins with Johnson and Greene’s Who Are Government Open Data Infomediaries? A Preliminary Scan and Classification of Open Data Users and Products. The authors sketch out the range of third-party infomediaries—classified as government, private sector, NGO, academic, and media—who access government open data and create value-added products. Classifying how infomediaries translate open data into actionable information enables, among other benefits, feedback to government open data providers looking to justify, improve, and expand open data initiatives.

The next paper, Exploring Open Data Perspectives from Government Providers in Western Canada, by Gill, Corbett, and Sieber, explicates the role of government personnel in open data production and provision in western Canadian jurisdictions of varying sizes—ranging from large city centers (Edmonton and Vancouver) to smaller cities (Victoria and Vernon) and regional districts (Okanagan). Through interviews of government open data personnel, paired with an assessment of supporting open data policies, the authors share an in-depth understanding of how the size of municipalities affects how open data initiatives develop.

The third paper in this issue, Open Data Community Maturity: Libraries as Civic Infomediaries, by Robinson and Maher, extends the scope of infomediaries to libraries as long-standing practitioners of information science. Libraries have played the role of data curators long before the introduction of digital information, and, as such, are ideally suited to serve as civic infomediaries capable of meaningfully connecting the nonexpert public with open data to create a more robust open data ecosystem.

Shifting towards the transformative potential of open data infomediaries, the fourth and fifth papers extend the conversation to nonhuman actors and explicitly address the role of data formats and software platforms in the access and use of open data. Baculi, Fast, and Rinner investigate The Geospatial Contents of Municipal and Regional Open Data Catalogs in Canada, a survey of the geospatial content of government open data and the platforms that enable their use. The focus on the availability and growth of geospatial and GIS-ready datasets in municipal open data catalogs, with a prevalence reported around the well-known 80 percent mark, highlights that data format and delivery method impact the potential uses of open data.

Lastly, Sangiambut and Sieber conclude this special issue with The Civic Open Data and Crowdsourcing App Ecosystem: Actors, Materials, and Interventions. They extend the conversation to the role human and nonhuman infomediaries play in the valuation and transformation of open data, with a focus on the complex network of mutual interdependencies between open data infomediaries and the civic-data app ecosystem. The authors caution that the inclusion of infomediaries in an era of crowdsourcing, open data, and Web 2.0 can reduce the role of government to a mere platform for service delivery.

The emerging and diverse roles of infomediaries explored in this issue support the making of open data. This work sets the foundation for necessary and immediate next steps to tackle more complex questions related to the critical evaluation of the value of open data and its limitations: The capacity to facilitate the interaction between and among governments and citizens; the perceived enhancements to transparency, accountability, and access; and the mechanisms that support or prevent input to open government.

This special issue is the culmination of interconnected research projects within the Canadian “Geothink” research partnership, a network of researchers funded by the Social Sciences and Humanities Research Council of Canada (SSHRC). The partnership examines the influence of the geospatial Web on reshaping government-citizen interactions. During the course of this project,
research on the geospatial Web has expanded into open data as one of its core areas of interest—a reflection of the societal trend towards open data. The research activities from Geothink are informed by close collaborations with professionals from federal, provincial, and municipal governments across Canada, private-sector firms, and nonprofit organizations. These collaborations have given Geothink researchers an in-depth and behind-the-scenes understanding of what it takes to make data “open.”

About the Authors

Dr. Victoria Fast is an assistant professor in the Department of Geography at the University of Calgary. As an urban GIS specialist, she applies GIS and data science, Web mapping, and participatory methods to engage government, citizens, and NGOs on issues related to building smart—accessible, livable, resilient, and sustainable—cities.

Corresponding Address:
Department of Geography
University of Calgary
2500 University Dr. NW
Calgary, Alberta, Canada
victoria.fast@ucalgary.ca

Dr. Claus Rinner is a professor and chair of the Department of Geography and Environmental Studies at Ryerson University. His research expertise within geographic information science covers geovisualization, participatory GIS, and spatial decision support. More recently, he became interested in open geospatial data, as exemplified in this special issue, and in 3-D–printed geographies.

Acknowledgments

This special issue emerged from research collaborations within a Social Sciences and Humanities Research Council of Canada (SSHRC) Partnership Grant 895-2012-1023, the “Geothink” project, and was also partially funded by a SSHRC Doctoral Scholarship. It was truly inspiring to work together to create this collection, and we thank the individual articles’ authors, reviewers, and the journal editors for their dedication, patience, and wisdom.
INTRODUCTION

In recent years, many governments across North America have begun to provide open data. This release of government information to the public, with minimal use restrictions, is used to enhance transparency and accountability of government, and to drive creative reuse of government data (Bedini et al. 2014; Janssen, Charalabidis, and Zuiderwijk 2012). Despite these accessibility-related motivations for data release, much of the open data currently published by governments is provided in a specialist format, annotated with government-specific terms, tags, or metadata, and often provided in “bulk,” covering a wide time frame or geographic area (Conradie and Choenni 2014, Sieber and Johnson 2015). This method of open data provision reflects the still-emerging mandate of governments as data providers to the public, with limited focus on the technical issues that such users may encounter. Rather, open data provision is based around government exposing data that often is found in the same format and structure as would be used internally. As a result, it can be challenging for nontechnical end-users to access, analyze, and apply open data to projects (Janssen and Zuiderwijk 2014; Zuiderwijk, Janssen, and Dwivedi 2015). This disconnect between the data creation and use by a specific set of users inside of government and its current sharing to a wider, more diverse user base outside of government can potentially limit the overall value of open data. Converting government open data into accessible information that can lead to action can necessitate a third party—the infomediary. An infomediary is an entity that takes open government data, in whatever format it may be provided in, and makes it more accessible and useful for end-users, through added analysis, combination with other data sources, and through visualization or publication (Janssen and Zuiderwijk 2014). For example, journalists act as infomediaries when they access government open data to write a story for publication. Also, community groups, NGOs, and private companies can act in this way, taking raw data, adding value through analysis, and then communicating these results to a specific audience. In this way, the value of open data is not limited to its simple provision, but rather in how use is made of it (Bruin, Bregt, and Ven 2001; Janssen et al. 2012).

A robust knowledge of infomediaries and their role in translating open data to actionable information is only just now emerging from academic literature, including a characterization of the main types of infomediaries, how they transform government data, and the uptake of services that they provide for others further along the chain of data use. The goal of this research is to present an overview of the types of infomediaries that access government open data. These categories of infomediaries include government agencies, the private sector, nongovernmental organizations and community groups, academics and researchers, and the media. This paper draws on an information scan of four leading Canadian municipal-government open data portals (Toronto, Edmonton, Vancouver, Ottawa) to identify the publicly available work of open data infomediaries. A classification of the main types and form of open data infomediaries then is created. We use this classification to advance a discussion on the role that infomediaries play in enabling broader access to and enhancement of government open data. This research provides a foundation for further research on the role and functioning of open data infomediaries, and can provide feedback to government open data providers looking to develop their open data user ecosystems.
OPEN DATA AND OPEN DATA INFOMEDIARIES

The provision of open data by government has increased dramatically in recent years. Open data is government data on infrastructure, spending, services, and procedures, made publicly available for free, online, under a permissive-use license (Bedini et al. 2014, Bonina 2013). Government open data portals are an important first step towards achieving some of the promise of open government by sharing data within and across governments, as well as with a broad community of users, including citizens (Bedini et al. 2014; Charalabidis, Alexopoulos, and Loukis 2016). The traditional open data portal model provides open data, online, in a variety of raw data formats. This “data over the wall” model is not necessarily focused towards broad or easy access by average citizens (Sieber and Johnson 2015). This focus on developers or other technically minded users compared to citizens comes despite long-standing pressure for government to provide information in citizen-centric formats (Luna-Reyes, Bertot, and Mellouli 2014).

Infomediaries can be defined as specific categories of open data users who extract, aggregate, and transform data, altering it into a format that is seen as valuable, beneficial, and, most important, usable to the general public (Bonina 2013). Government infomediaries include open data providers as well as other government departments within a municipality that create value-added products with open data, making it more accessible to the end-user (Janssen and Zuiderwijk 2014; Magalhaes, Roseira, and Strover 2013). Infomediaries can play a variety of roles, ranging from increasing transparency and accountability to encouraging innovation and economic growth within a city (Davies and Frank 2013). Generally, an important role of a government infomediary is the ability to visualize complex relationships and information, creating value-added services for other departments within the government, and additionally for citizens (Deloitte Analytics 2012). Using tools for visualization and data analysis, such as the creation of charts, graphs, and applications, improves accessibility to open data for any audience, particularly those without a technical background (Fumega 2014, Graves and Hendler 2013). Technically skilled infomediaries are critical to the success of open data programs, for they have the ability to engage citizens who otherwise would not access open data because of a lack of interest, skills, or resources (Fumega 2014). Infomediaries act to help bridge the gap between the data that is available from government and the potential for it to be used by citizens. This provides an opportunity for a range of end-users to both access and effectively use this data (Chattapadhyay 2014, Davies 2014, Fumega 2014).

G4 CITIES AND METHOD

Open data provision in Canada has expanded dramatically over the past decade. Starting as an emerging phenomena at only a select few cities, open data now is commonly provided by many municipal and provincial governments. Many of those municipal governments offering open data do so using the Canadian Open Government License, developed by the federal Canadian government, which provides for attribution-only, mixed commercial/noncommercial use of open data (Roy 2014). In parallel to this increased provision of open data by governments, private-sector use of open data in Canada is growing. The recent launching of the Open Data 150 found more than 150 private-sector companies using Canadian open data for a variety of purposes (GovLab 2017). Given this increase in both provision and use of open data, we investigate the current range of infomediary activities in major Canadian open data—providing cities, using a targeted review of publicly available information.

Taking the “G4,” the most developed open data—providing Canadian cities of Edmonton, Toronto, Ottawa, and Vancouver as our study subjects, we conducted a Web scan of municipal Web sites, including searching their open data catalogs, available galleries of finished products, results from hackathons or other coding contests, and other departmental pages to gather examples of infomediaries. Data for this research was collected from September to December of 2015. Additionally, targeted key searches were undertaken to find examples of products created by a specific infomediary category in a G4 city. These targeted searches included some key terms such as Toronto AND open data along with more specific searches such as Vancouver AND data journalists. Results then were compiled to reveal both general results of the types of infomediary products created and the open data used, and, additionally, a profile of each infomediary type was created, showing commonalities and differences among the G4 cities. Though an exhaustive record of all open data use by infomediaries is not possible, this research gives a starting point for the classification of categories of infomediary, an overview of the types of products created from major municipal open data catalogs, and how these products are delivered to end-users.

CLASSIFICATION OF INFOMEDIARIES, INFOMEDIARY PRODUCTS, AND DELIVERY METHOD

Infomediary products for the G4 cities were classified in three ways: infomediary type, infomediary product, and delivery method. First, the classification of infomediary type drew from existing literature on open data—user communities, most notably Bonina’s (2013) overview of open data business models, Magalhaes et al.’s (2013) framework of open-government intermediaries, and Janssen and Zuiderwijk’s (2014) classification of infomediary business models. This literature, plus other works, describes key open data—user communities as including government itself (Kankanhalli, Zuiderwijk, and Tayi 2017), the private sector (Bonina 2013), journalists and media organizations, researchers and academics, and nonprofits or community groups (Chan, Johnson, and Shookner 2016; Safarov, Meijer, and Grimmelikhuijsen 2017). Given the identification of these user communities, we adopt a similar classification to frame this research, though we also acknowledge that a complete census of all open data users is difficult to track (Johnson 2016).
To frame this research, we used five main types of infomediaries: government, private sector, NGO and/or community groups, academic researchers, and journalists/media. Table 1 shows a breakdown of infomediary type by each of the cities included in this research. These broad categories serve as a high-level typology of the open data infomediary role. Infomediary product categories were created through aggregating similar-type products found at the surveyed cities. Table 2 presents these types of products by infomediary type. Though we classified ten different types of infomediary products, significant numbers are found only in four types: searchable database (such as a transit, recreation, or festival schedule), interactive map (using maps to display a user-selected range of information, such as municipal infrastructure), articles that contain data visualizations (prepared for mass media, blogs, etc., containing charts, tables, maps), and find-a-service portals, such as locating street food vendors or historic buildings. Lastly, delivery method of infomediary products was classified into one of four categories: database (a user-query dataset, but with limited visualization or analysis capabilities), a Web site (publicly posted document that may contain text, images, maps, and information derived from a database), mobile app (similar to a Web site but designed specifically for a telephone interface), and a specific computer application to run on a desktop or laptop computer. Delivery method was used to track the split between products provided in specialist formats, broadly on the Web, and as a mobile app, targeting mobile telephone users. Table 3 gives a breakdown of each of these delivery methods based on infomediary type, as well as links to examples. These results are presented within each infomediary type, across all surveyed cities, with breakdowns of type of product and delivery method. These latter two areas (products and delivery method) are described in context of each of the high-level infomediary types—government, private sector, NGO, academic, and media.

### Table 1. Breakdown of infomediary type by city

<table>
<thead>
<tr>
<th>Category</th>
<th>Edmonton</th>
<th>Ottawa</th>
<th>Toronto</th>
<th>Vancouver</th>
<th>Total</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government</td>
<td>29</td>
<td>13</td>
<td>12</td>
<td>8</td>
<td>62</td>
<td>27</td>
</tr>
<tr>
<td>Private Sector</td>
<td>18</td>
<td>49</td>
<td>33</td>
<td>26</td>
<td>126</td>
<td>55</td>
</tr>
<tr>
<td>NGO</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>4</td>
<td>14</td>
<td>6</td>
</tr>
<tr>
<td>Academic</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td>7</td>
<td>3</td>
</tr>
<tr>
<td>Media</td>
<td>5</td>
<td>0</td>
<td>11</td>
<td>3</td>
<td>19</td>
<td>8</td>
</tr>
<tr>
<td>Total</td>
<td>55</td>
<td>68</td>
<td>63</td>
<td>42</td>
<td>228</td>
<td>100</td>
</tr>
<tr>
<td>Percent</td>
<td>24</td>
<td>30</td>
<td>28</td>
<td>18</td>
<td>100</td>
<td></td>
</tr>
</tbody>
</table>

### Table 2. Type of product by infomediary type

<table>
<thead>
<tr>
<th>Product Type</th>
<th>Gov %</th>
<th>Private Sector %</th>
<th>NGO %</th>
<th>Media %</th>
<th>Academic %</th>
<th>Total %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Searchable Database</td>
<td>20</td>
<td>64</td>
<td>51</td>
<td>7</td>
<td>7</td>
<td>91</td>
</tr>
<tr>
<td>Interactive Map</td>
<td>23</td>
<td>40</td>
<td>32</td>
<td>1</td>
<td>7</td>
<td>29</td>
</tr>
<tr>
<td>Reporting Service</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>Find a Service</td>
<td>2</td>
<td>15</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>17</td>
</tr>
<tr>
<td>Report</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>7</td>
<td>14</td>
</tr>
<tr>
<td>Static Map</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>7</td>
<td>9</td>
</tr>
<tr>
<td>Graph</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>14</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>Game</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Web Page</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Articles</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>14</td>
<td>3</td>
</tr>
<tr>
<td>Total</td>
<td>62</td>
<td>126</td>
<td>100</td>
<td>14</td>
<td>19</td>
<td>228</td>
</tr>
<tr>
<td>%</td>
<td>27</td>
<td>55</td>
<td>6</td>
<td>8</td>
<td>3</td>
<td>100</td>
</tr>
</tbody>
</table>

To frame this research, we used five main types of infomediaries: government, private sector, NGO and/or community groups, academic researchers, and journalists/media. Table 1 shows a breakdown of infomediary type by each of the cities included in this research. These broad categories serve as a high-level typology of the open data infomediary role. Infomediary product categories were created through aggregating similar-type products found at the surveyed cities. Table 2 presents these types of products by infomediary type. Though we classified ten different types of infomediary products, significant numbers are found only in four types: searchable database (such as a transit, recreation, or festival schedule), interactive map (using maps to display a user-selected range of information, such as municipal infrastructure), articles that contain data visualizations (prepared for mass media, blogs, etc., containing charts, tables, maps), and find-a-service portals, such as locating street food vendors or historic buildings. Lastly, delivery method of infomediary products was classified into one of four categories: database (a user-query dataset, but with limited visualization or analysis capabilities), a Web site (publicly posted document that may contain text, images, maps, and information derived from a database), mobile app (similar to a Web site but designed specifically for a telephone interface), and a specific computer application to run on a desktop or laptop computer. Delivery method was used to track the split between products provided in specialist formats, broadly on the Web, and as a mobile app, targeting mobile telephone users. Table 3 gives a breakdown of each of these delivery methods based on infomediary type, as well as links to examples. These results are presented within each infomediary type, across all surveyed cities, with breakdowns of type of product and delivery method. These latter two areas (products and delivery method) are described in context of each of the high-level infomediary types—government, private sector, NGO, academic, and media.
GOVERNMENT INFOMEDIARIES

Government infomediaries include open data providers themselves, governments at the same or alternate level (municipal, state/province, federal), as well as departments within a given government that create different value-added products using government open data (Jansen and Zuidervijk 2014; Jetzek, Avital, and Bjørn-Andersen 2013). From our study sample, 27 percent of all the examples of infomediary activity gathered across the G4 cities were produced by government infomediaries (see Table 1). This was the second largest category of infomediary, after the private sector, though we discuss this infomediary first, for it is the closest to the originating source of data. With the delivery of open data providing a foundation for further actions, other government departments or agencies act as infomediaries through the analysis of data and resulting production of more accessible content directed towards a specific range of end-users (Deloitte Analytics 2012). This value-added creation can take the form of visualization or data analysis, such as the creation of interactive maps, charts, graphs, and applications that improve the accessibility to open data for many audiences, particularly those without a technical background (Fumega 2014, Graves and Hendler 2013).

Table 2 provides a breakdown of the main types of products made by government infomediaries. For example, across all G4 cities, the use of interactive maps (37 percent) was a common way to allow users to display, filter, and perform basic analysis. Searchable databases (32 percent), such as a listing of recreation programs or transit times, also were commonly seen types of products. Additionally, the creation of static maps (10 percent) highlighting such city services as bicycle routes was seen, as well, as a less common but still consistent form of sharing open data information. Creating visualizations for reports (8 percent) through specialized data-management software, such as Tableau (http://www.tableau.com/) also were seen throughout the G4 cities, particularly for sharing complex budget information through reports that relied heavily on simple graphics including charts and graphs. An example of this is the city of Edmonton’s Citizen Dashboard (https://dashboard.edmonton.ca/), an online tool designed to engage nontechnical end-users through easier-to-understand visualizations of municipal performance indicators. These infomediary products were delivered largely as Web sites (75 percent) and as mobile apps (18 percent), showing government infomediaries as focusing on wider distribution via the Web, compared to more fragmented mobile operating systems (see Table 3). Overall, across the G4 cities, government infomediaries were critical players in working with government open data to create value-added products.

PRIVATE SECTOR

The private sector is considered to drive demand for open data through the potential for economic growth and overall ability to innovate and create services for citizens (Bonina 2013, Davies 2014, Gray and Darbishire 2011). Private-sector infomediaries include companies and individual developers who use open data to create new products (Deloitte Analytics 2012). In particular, the creation of mobile applications has been seen as a major outlet in which the products of open data can be shared widely (Yang and Kankanahalli 2013). Across the study sample, private-sector companies represented 55 percent of all infomediaries found in the G4 cities (shown in Table 1). This is the largest category of infomediary, with more than twice as many examples as the next largest category (government).

Previous research has found that open data is used by private-sector developers to design Web sites and maps and build Web-based or mobile applications, which may include real-time information (Davies and Edwards 2012, Deloitte Analytics 2012) into research, to operational project statistics. A common example of a product created by private-sector developers is a public-transit application. These applications use real-time bus Global Positioning System (GPS) information to help citizens interact more efficiently with public transit (Rojas 2012). The sample of open data products created by the private sector using G4 city data is dominated by three main types of products: searchable databases (51 percent), interactive maps (32 percent), and find-a-service (12 percent) (see Table 2). Searchable databases created by private-sector infomediaries included transit routing apps. With bus schedule data available to the public in all G4 cities, many developers have used this data to create transit applications that show bus routes and allow for trip planning. Interactive maps created by private-sector infomediaries include real-estate applications that allow users to filter demographic and neighborhood amenity information along with for-sale listings. Find-a-service applications include third-party apps that facilitate interaction with data such as the location of public-access defibrillators (Ottawa) or festival events (Toronto). Overall, the types of products created by private-sector infomediaries is very diverse. Another notable product created by a private-sector infomediary includes the samples-only game created using open data of the cities of Ottawa, Toronto, and Vancouver. This Web-based interactive game titled “Click that ‘hood” (http://click-that-hood.com/) engages citizens in learning the names of all the neighborhoods located within a city. Although this was the only game found in the research, it shows the potential of open data to be used outside of the realm of just informing or guiding citizens through a city, but additionally engaging citizens in enjoyable activities.

Private-sector infomediaries primarily delivered their open data–based products as mobile applications (62 percent, shown in Table 3). Many of these applications were creations from hackathon events, hosted by the G4 cities. The private-sector companies creating mobile applications therefore ranged from one-off applications by individual developers to a number of applications made by larger public companies. This significant use of the mobile application for delivering open data products shows the integration of private-sector infomediaries with the more easily monetized mobile-app ecosystem, compared to a Web site or other method.
NONGOVERNMENTAL ORGANIZATIONS (NGO)/COMMUNITY GROUPS

Nongovernmental organizations (NGOs) and community groups are infomediaries who produce a wide range of products with open-government data (Chattapadhyay 2014). An important role played by this type of infomediary includes encouraging government accountability and effectiveness (Davies 2014). NGOs provide key insights into which datasets should be opened to help produce high value and engagement with the public (Ubaldi 2013). Another task performed by NGOs includes the creation of reports, most of which focus on highlighting education, health, and other societal issues (Graves and Hendler 2013). NGOs additionally focus on sharing budget-related information in easy-to-understand formats (Mejabi, Azeez, Adedoyin, and Oloyede 2014). An important function of NGO and community-group infomediaries are their ability to collaborate with governments, building trust in opening information to the public, and additionally opening data to a wider audience of potential users, including grassroots communities (Gray and Darbishire 2011). Although NGOs are seen as a smaller group of infomediaries for open data, they are growing in number and importance through their engagement with open data in relation to domestic priorities and issues (Davies 2014).

NGO and community-group infomediaries represented 6 percent of the total number of all infomediary products made with G4 city data (shown in Table 1). Additionally, for NGO and community-group infomediaries, the sample size is very small, with only 14 individual products recorded across the four cities surveyed. Infomediary products produced by NGOs in the G4 cities were dominated by searchable databases (50 percent), with only a handful of singular instances of other categories. Delivery methods used to provide these products were dominated by Web sites (57 percent), a sharp contrast to the mobile app-heavy provision of the private-sector infomediaries (see Table 3). This contrasts to literature on the role of NGOs and nonprofits in accessing and using open data to create products of wider value to citizens. Despite low individual numbers of infomediary products, there were notable examples that could have broad use by citizens and others. For example, in the city of Vancouver, a collaborative project was created called Urban Opus (http://urbanopus.net/), an organization that brings together community organizations, governments, and private-sector corporations that helps to develop original data-driven services to engage civic audiences. Similarly, Geographic and Effective Measures Services (http://www.gems-spc.ca/) is another consulting NGO located in Ottawa that creates profiles of given neighborhoods, as well as custom printing maps. Lastly, the rePresent API, a project of Open North (https://represent.opennorth.ca/), uses a range of open data to provide a platform for informing citizens about their local political representatives.

Academics/Research

Research infomediaries include academics and consultants that use open data to create academic papers or reports, analyzing topics related to government open data (Davies 2010). This group of infomediaries analyzes open data and provides the results of the analysis through both text and visualizations. The purpose of research infomediaries is to use open data to solve important issues and provide insights into government actions and services. Commonly, transit ridership was analyzed to determine the success of transit systems in different municipalities. Open data additionally can be used alongside other datasets to enhance research and potential results. Research infomediaries are of particular importance for their products include value added to the original data, providing further information and analysis of governments to citizens. Research infomediaries in many cases perform secondary research, using government open data to cross-check findings, provide new findings, and generally inform citizens about government-related information (Davies and Edwards 2012). Across the G4 cities sample, research infomediaries were found to represent only a small number of the total number of infomediary examples (3 percent, see Table 1). Given this small total number of examples, it is difficult to draw any meaningful conclusions about the specific types of infomediary products created, though of those examples, 57 percent were articles and 29 percent were interactive maps (shown in Table 2). It is notable that all the research infomediary products were delivered using the Web page method (see Table 3).

Table 3. Product delivery method by infomediary type

<table>
<thead>
<tr>
<th>Infomediary Type</th>
<th>Gov %</th>
<th>Private Sector %</th>
<th>NGO %</th>
<th>Media %</th>
<th>Academic %</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Database</td>
<td>5</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Web Site</td>
<td>46</td>
<td>74</td>
<td>42</td>
<td>33</td>
<td>8</td>
<td>57</td>
</tr>
<tr>
<td>Mobile App</td>
<td>11</td>
<td>18</td>
<td>78</td>
<td>62</td>
<td>3</td>
<td>21</td>
</tr>
<tr>
<td>Computer App</td>
<td>0</td>
<td>6</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>62</td>
<td>100</td>
<td>126</td>
<td>100</td>
<td>14</td>
<td>100</td>
</tr>
</tbody>
</table>
Media

Primarily titled data journalists, media infomediaries access open data, analyze the information, and provide citizens with results and analysis of the information. With the increase in social media, more and more venues are opening up in which data journalists can share reports on open data, allowing citizens to discuss and comment on the information provided. The introduction of wider structures of media networks, such as social media, has positively impacted the availability and sharing of open data to citizens (Davies 2010). With the rise of open data, there has been a creation of a new group of journalists known as data journalists. Data journalists act as infomediaries by accessing government open datasets to identify stories and share information with the general public (Mejabi et al. 2014, Roy 2014, Sapkota 2014). The role of data journalists is seen as fundamental in distributing government data and engaging citizens in government open data (Furnega 2014). Particularly in developing countries and/or communities, data journalists are seen as important infomediaries for they are most accessible to citizens, including those who may not have access to or understand technologies such as the Internet (Mejabi et al. 2014, Sapkota 2014). Across the G4 cities sample, media represented a small total number of instances of infomediary products, with 8 percent of the total amount (see Table 1). Despite this small number of specific instances, there is potential for a large audience or readership with media infomediary products. Overall, media infomediaries created articles, with 100 percent of the infomediary products being articles (see Table 2). These articles were uniformly distributed via Web sites (shown in Table 3). Blog posts were a common form of journalism including government open data. Many of these blog posts include tutorials on how to use and access open data, teaching citizens about the data they are using.

A CLASSIFICATION OF OPENDATA INFOMEDIARIES AND “WEBS” OF OPEN DATA VALUE

Through this characterization of the Canadian G4 cities open data infomediaries, several key themes have emerged. First is that there is a diverse range of organizations, whether government agencies, private-sector companies, media, researchers, or non-profits and community groups that access government open data. These broad categories of infomediary each access data for a variety of purposes and use open data in various ways related to their organizational and individual goals. Despite this broad range of actors that access open data, the infomediary landscape is dominated by private-sector and government infomediaries. As characterized, the types of products created through the use of open data by infomediaries are varied and can include databases, Web sites, mobile apps, blog posts, reports, and more. Table 2 provides an overview of these uses of open data and also the ways in which they are distributed, either through Web sites, mobile apps, or dedicated conventional computer programs. Given this range and diversity of intervening actors and end-products created, a significant outcome of this work is to suggest a refocusing of current attention on open data from the simple provision of open data to the questions of how other entities are accessing this data to create downstream value. The identification of specific users, each with unique use cases, can be used to better understand the impact of data provision by government. Within open data literature and practice, there is a significant focus on enabling access to raw data from government (Denis and Goëta 2014, Sieber and Johnson 2015). This literature focuses largely on topics of data format, licensing and copyright, as well as organizational issues that may impede government provision of open data (Chattapadhyay 2014, Harvey 2007). These concerns are typical of early innovation life-cycle issues (Rogers 2010) where technical or mechanical issues, those of just providing data (Johnson and Sieber 2013, Sieber and Johnson 2015), are paramount. As open data matures and becomes more standardized across governments, this discussion moves naturally on to one of value and utility of the data access provided. It is to this growing field of research that this work aims to contribute, through the classification of the infomediary class of open data users.

Building on the classification of open data infomediaries presented here, we propose that infomediaries are a primary way that open data creates a “web” of value, with a single point of access by an infomediary creating broader access through value-added developments driven by third parties. This concept is supported through recent studies from Janssen (2012) and the New York University GovLab (Verhulst and Young 2016), both of which aim to evaluate the impact of open data. Janssen (2012) indicates that open data itself has little value, and that this value is only realized after use. Publication of data in an open format for easy access is simply the first step towards use and impact; however, for this impact to be generated, infomediary action is required. Impact of open data is considered by GovLab (Verhulst and Young 2016) as the product of a series of enabling conditions, including the leveraging of open data by partnerships both within and external to the data-generating government. This presentation of partnerships as a key to unlocking the value of open data is a finding that is revealed in this work as well, with the infomediary class of open data users representing a manifestation of the partnership or open contract (whether explicit or implicit) between the data-generating government and the user (Verhulst and Young 2016). This contract includes not only the ability to access data, but to enhance, manipulate, and, in turn, share open data to other downstream users, forming the “web” of value where impact, generated through initial sharing between government and infomediary, is enhanced through infomediary activity. For example, in the survey of G4 cities, infomediaries often combined datasets (both open and closed) and enhanced open datasets to create a value-added proposition for downstream users. Through the sharing of data in this fashion, impact also is diversified, moving into areas that may not have been initially considered by the data owners. An example of this is the aforementioned “Click that ‘hood’” game, where the creative use of neighborhood boundary
data has potential to increase civic pride and knowledge of an area, an outcome that is vastly different than the likely intended goal of providing neighborhood boundary data to facilitate record keeping, map making, and other typical uses.

Despite the potential formation of webs of open data value by infomediary actors, many questions remain. First, as revealed by this research, products created by government and private-sector infomediaries vastly outnumber those created by other infomediary groups. This may indicate that these two categories of infomediary are the most significant users of government open data; however, this may ignore the extent to which these products actually are used. For example, though the private sector may create a number of products from government open data, how many of these products are essentially duplicates of one another, with minor differences? And how many of these products generate a robust number of users? These relevant questions indicate that in many ways, tracking the value of open data requires the tracing of all nodes on the data access and use web, following through from transformations, analyses, or unique data combinations that may be introduced by infomediaries. Specific numbers or quantitative measures of use, such as downloads of datasets, number of products created, or even the number of viewers of a particular news story partially made with open data, while straightforward to quantify, lack in understanding of the impact and role played by open data, particularly in further downstream uses of products created by primary infomediary activity. Though a finding of this research is that the open data infomediary space is dominated by government and private-sector actors, this should not be extended to include that these infomediaries necessarily lead to a greater impact of open data via these activities.

CONCLUSIONS

This research provides a characterization of open data infomediaries, using four major Canadian cities as sources for comparing the types of infomediaries, products produced, and delivery methods used. Infomediaries were framed into five distinct categories, based on previous research and sectoral characteristics: government, the private sector, NGO or community organizations, academics or researchers, and the media. The total number of instances of infomediary products across the four cities was dominated by the private-sector and government categories of infomediary. Products produced by infomediaries depended largely on the type of infomediary, but typically focused on searchable databases, interactive maps, find-a-service tools, and media/blog articles. Delivery methods were dominated by Web sites and mobile applications. A significant message from this work for governments that provide open data is that enabling access to data, though clearly important, is only the first step in generating impact from open data. Providing access to data is simply the beginning of what can be considered a “web” of open data use that is driven by infomediary actors. As noted by Janssen et al. (2012), the simple provision of open data has little value. It is from the process of use, via numerous infomediaries, that value is created.

It is important to consider that all open data programs are not identical, and that the individual model of open data provision may support or restrict infomediary activity, or even favor certain types of infomediaries. For example, Sieber and Johnson’s (2015) models of open data provision indicate some forms of data provision would better support private-sector infomediary activity, most notably the “data over the wall” model. Other models, such as “code exchange” and “civic issue tracker,” more deeply involve government in creating data endpoints for citizen use of open data (Sieber and Johnson 2015). Lastly, the prospect of participatory open data is a model where citizens and government co-create products with open data. This model would firmly cement the role of government as an infomediary when dealing with citizens, with deep responsibility not only for the data itself, but in how it gets used.

Many challenges exist to the use of open data, and infomediaries accessing and using data many encounter many technical, organizational, and social challenges. Technical challenges include differences in data collection, format, licensing, and varying levels of data completeness and quality that may impede use (Johnson, Sieber, Scassa, Stephens, and Robinson 2017). For government to support the use of open data by infomediaries, there are many changes to data provision that could ameliorate or reduce these challenges. First, the development and enforcement of open data standards (Goëta 2014, Piovesan 2015) can provide infomediaries with more seamless data that crosses jurisdictions. This can support the use and integration of multiple datasets from across different governments. An agreed-upon set of standards for the provision of open data also can ensure that infomediaries can access data in a uniform fashion, reducing development costs generated through working with and processing data in a variety of formats.

At a higher level, the provision of open data by governments to a range of infomediaries, some of which may be building billion-dollar business models using open data, raises a number of questions. First, what is the role of government in supporting, for free, these types of corporate endeavors (Johnson et al. 2017)? From the business-development side, can businesses rely on government to continue providing, collecting, and licensing data for free, or could cost-recovery fees be reintroduced to support government data-collection and management efforts (Sieber and Johnson 2015)? Lastly, what priority will government place on the development and provision of data that meets a corporate need versus data that meets a societal need? Could a priority be placed on that data that is mission-critical for private-sector infomediaries (with teams of lawyers and effective lobbyists), compared to data that may have little commercial value, but instead be of broader social value, or political value, such as transparency or procedural data? These are critical next-step questions that will further refine the role of the infomediary and its evolving relationship with the data-generating government.

Ultimately, the development of a class of open data infomediaries indicates that data-provisioning governments have many different user communities to work with, and that there may be challenges in meeting the potentially divergent needs of
these communities. A particularly notable challenge is balancing the technical needs of developers versus the nontechnical needs of more general users. With the development of infomediaries, government no longer can simply provide data, but rather must be responsive to the needs of infomediaries who may request standardization of formats, specific datasets required to meet their needs, and further support from the data custodians. This issue, of access to technology by those with varying levels of skill to use a specific piece of software, has been noted in early work on diffusion of GIS (Onsrud and Pinto 1991) and the development of public participation GIS (PPGIS) approaches (Sieber 2006). In contrast with these works, which largely looked at access on an individual or group level, the role of an infomediary organization can be to overcome some of these barriers to access of technology and data that are encountered by society at large. Though the specific functioning of the infomediary as an open data “chauffeur,” helping to spread access and skills in open data analysis and application, is beyond the scope of this work, it remains an important line of questioning, and one that can draw on GIS implementation literature more strongly (Ghose 2001, Sieber 2001).

This research presents an initial view of the role of infomediaries in generating value from government open data within a convenience sample of cities—the G4 Canadian open data cities of Toronto, Vancouver, Edmonton, and Ottawa. As this represents a small subset of all open data–providing cities, this work should be considered exploratory. Additionally, as the total number of infomediaries using government open data cannot entirely be known, there is the possibility that the relative breakdowns of open data infomediary type and product presented here are a preliminary view. Given the emerging state of research on the nature and output of open data users, there still are important lessons to guide future research. Most notably, that government is a major user of open data shows that the infomediary role is broad, and that government has a role to play in open data use, not just provision.

The benefits of open data may take time to emerge, as private-sector users usually look for market opportunities rather than observing the open data to invent new products immediately (World Wide Web Foundation 2015). Although the impacts and benefits of open data may not be present at first, defining them is important in discovering the importance of open data programs (Höchtl, Davies, Janssen, and Schieferdecker 2014). Understanding the relationship between those who publish the data, those who create products based on the open information, and the final end-user is where many of the overall impacts can be found. As many citizens themselves may not be able to interact directly with open data, the role of infomediaries becomes important in determining the overall success of an open data program. Now that many of the initial challenges in providing open data have been surmounted, the development of measures of open data value are taking priority in the research and practitioner fields. As open data moves from a disruptive phenomenon to a commonly found, and in many ways, expected service, there is increased pressure to justify the continued expense and effort of data provision. In this way, the development of metrics of value are a key component to creating a picture of open data use that includes diverse end-users and also those infomediaries who reach a wider diversity and extent of end-users.

Acknowledgments

This work was supported by the Social Sciences and Humanities Research Council of Canada (SSHRC) Partnership Grant Geothink.

About the Authors

Dr. Peter A. Johnson is an 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, including open data, the geospatial Web, mobile devices, crowdsourcing, participatory geomatics, and volunteered geographic information.

Corresponding Address:
Department of Geography and Environmental Management
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada
peter.johnson@uwaterloo.ca

Sarah Greene completed her undergraduate degree in geography and environmental management at the University of Waterloo in 2016. She currently is attending Ryerson University, completing a Master’s of Spatial Analysis. Her research interests include studying the impacts and use of open data.

Corresponding Address:
Department of Geography and Environmental Management
University of Waterloo
200 University Avenue West
Waterloo, Ontario, Canada N2L 3G1
sagreene@uwaterloo.ca
References


Exploring Open Data Perspectives from Government Providers in Western Canada

Mark Gill, Jon Corbett, and Renée Sieber

Abstract: This paper explores the roles that municipal government open data personnel play in the open data movement, and how their emerging roles can increase the societal and economic benefits of open data as well as increase citizen engagement. It then demonstrates how these roles are affected by the “scale effects” of the municipality. By interviewing municipal government open data personnel in western Canada and assessing current open data policies, we look at how open data initiatives are developing and identify where gaps can occur between governments and citizens from the government perspective. Finally, we look at how the size of municipalities affects how open data is developing at different-size municipalities. Currently, there has been limited research into municipal open data initiatives that take into account differences in size. We discuss three issues to emerge from this research that represent the way that government open data initiatives shape the interaction between governments and citizens. These include not knowing who open data users are, the jurisdictional nature of the way that data is managed, and the way that governments perceive what constitutes open data. We hope that by understanding how size is implicated in open data, governments will be more effective in their initiatives and provide better open data services.

INTRODUCTION

In a society that increasingly relies on data and technology, local and regional governments are making efforts to open up public-sector data to their citizens. Open data is data that is made available to the public that can be “freely accessed, used, modified, and shared for any purpose” (opendefinition.org). The open data movement is driven by numerous factors, including access to governmental data, improved service delivery, civic engagement, and participation in government (Sieber and Johnson 2015). Many governments have responded to this movement by making efforts to become more transparent and accessible to their citizens. One method to accomplish this is by making their data more accessible in an open data format; however, many improvements still can be made to close the data gaps that are formed between citizens and government. As Bates (2014) notes, there has been very little critical exploration of open-government data in relation to the agenda of government’s release of particular datasets and its relationship to economic, social, and political contexts. Many researchers have theorized about the potential benefits of open data (Jaeger and Bertot 2010, Restakis et al. 2014); however, particularly at the municipal level, there has been little research that explores how these benefits align with real open data initiatives.

Currently, access to open data has largely remained limited to government Web sites. Although the open data movement is meant to increase access to information, critics note a number of limiting factors, such as poor Web site design, restricted search functionality, limited opportunity for civic participation, and loose standards for the type and quality of data available (Jaeger and Bertot 2010, Janssen et al. 2012, Pina et al. 2007). Despite the potential for open data, assessments of municipal initiatives in Canada (Currie 2013, Davies and Lithwick 2010) have primarily focused on major, city-scale municipalities. There is limited research on open data within smaller municipalities, who now are implementing their own ambitious programs. In this sense, although there has been significant work on Canadian civic open data (for example, Hamilton 2015, Jofre et al. 2016, Sieber and Johnson 2015, Sieber et al. 2015, Robinson and Johnson 2016), it has not included scale as a comparative factor of both large and small municipalities. Given that more than six million Canadians live in centers of fewer than 100,000 (Statistics Canada 2011), we feel that this provides an important insight into open data initiatives at differing scales that have not previously been studied.

This paper explores the roles that municipal government open data personnel play in the open data movement, and how their emerging roles can increase the societal and economic benefits of open data as well as increase citizen engagement. It then compares how these roles are affected by the perceived scale of the municipality. By talking with municipal government open data personnel in western Canada, and assessing current open data policies, we look at how open data initiatives are developing as well as identifying where gaps occur between governments and citizens from the government perspective, and look at how the sizes of municipalities affect perceptions of open data. Currently, there has been limited research into municipal open data initiatives that take into account differences in size. We discuss three issues to emerge from this research that represent the way that government open data initiatives shape the interaction between governments and citizens. These include not knowing who open data users are, the jurisdictional nature of the way that data is managed, and the way that governments perceive what constitutes open data. We hope that by understanding how size is implicated in open data access and usage, governments will be more effective in their initiatives and provide better open data services.
LITERATURE REVIEW

Open government is a movement that has come about as a way to increase access to information, governmental transparency and accountability, and citizen/government engagement (Bates 2014, Sieber and Johnson 2015). The logic is that by increasing citizens’ access to government information, including raw data, the civic body will be able to monitor governmental services, policies, and procedures, to make sure that governments are accountable and transparent. In addition, through new mechanisms, citizens also will be able to better provide input on governmental affairs. Often open data initiatives are embedded within larger open-government initiatives that seek to increase civic participation and foster governmental transparency and accountability. Fittingly, research has shown that open data can act as a mechanism to enhance governmental transparency, civic participation, and economic innovation (Jaeger and Bertot 2010, Restakis et al. 2014). The Canadian government has stated that open data has the potential to “enable citizens, the private sector, and nongovernment organizations to leverage it in innovative and value-added ways” (Government of Canada 2014a).

In 2011, Canada joined the International Open Government Partnership (www.opengovpartnership.org). As a member, the government of Canada has established an Action Plan on Open Government designed to develop Canada’s open data initiative underneath a broader set of ideals known as “open government.” Although the federal government is adopting an open data initiative, provincial and municipal governments are responsible for their own data, and much that has been accomplished at the federal level has not yet reached lower levels of government. According to the government of Canada, an open government is one that seeks “to broaden access to data and information, ensure transparency and accountability, and strengthen citizen engagement in the activities of government and in the democratic process” (Government of Canada 2014b, 1). In 2012, Canada’s Action Plan on Open Government contained commitments to increase access to information, civic participation, and data innovation by trying to make data “open by default.” These commitments aimed to create policy around releasing data efficiently and automatically, while making the data easy to find, use, and understand. Canada’s Action Plan is the main mechanism for organizing the way in which open government, open data, and civic engagement is being addressed. Canada’s first Action Plan on Open Government (completed between 2012–2014) saw the introduction of Canada’s open data portal (data.gc.ca); access to information services (ATI) was digitally streamlined; a single open-government license for all data was issued; a new government-wide Web site (canada.ca) was launched; and Canadian open data experience (CODE), a data hackathon (a public event where citizens are encouraged to use open data to solve problems), was held in February of 2014 (Government of Canada 2014b). The most recent iteration of this Action Plan, released in 2016, which focuses on openness and transparency, has four broad themes: Open by default; fiscal transparency; innovation, prosperity and sustainable development; and engaging Canadians and the world. Scassa (2016) notes that the newest iteration of Canada’s Action Plan on Open Government has stressed transparency and openness because of the criticism it received for previous plans not focusing enough on these issues.

Although Canada’s Action Plan suggests an open and more citizen-inclusive process to governance, Doberstein (2013) notes that the Canadian government has held onto a more traditional, top-down style of governance while still trying to accommodate a more participatory approach. This balancing act between tradition and participation is termed metagovernance. Metagovernance is a form of “network governance” that Doberstein defines as the “increasing willingness by the state to engage with civil society partners on policy and development and implementation, while maintaining traditional Westminster systems of accountability and democratic legitimacy” (2013, 587). In this sense, citizen involvement in policy is a supplement to government, not a replacement (Woodford and Preston 2011). The key difference between metagovernance and standard network governance is that metagovernance maintains an overarching control over the decision-making process and outcome, while network governance is typically more horizontal, inclusive, and has a more egalitarian power structure (Doberstein 2013). According to Doberstein, metagovernance is extensively used at the local level in Canada but not at the provincial or federal level where critics have noted the lack of meaningful horizontal governance (Haddow and Klaassen 2006, Skogstad 2003).

The trouble that Doberstein sees in the many variations of metagovernance in Canada is that too often it is used as a “non-serious attempt to incorporate diverse views” (2013, 592); the input from citizens and civic bodies have little effect on policy making. An example that Doberstein uses is the role of an advisory board such as the District Health Councils in Ontario. Instead of having any policy-making power, the councils were used as a political buffer between local health institutions and the provincial government. The public saw that the recommendations of the District Health Councils were not being taken into serious consideration by the province’s policy makers. The dissatisfaction with the District Health Councils led to the formation of the Local Health Integration Networks (LHINs). These have real decision-making power over policy, funding, and coordination of services. When there is empowerment, local administrators and citizens are willing to engage with metagovernance networks because they feel their input actually can make a difference (Doberstein 2013).

Although there is potential for open data to have these effects, some research suggests that in its current state, open data does not meet that potential and instead is based in a “customer-centric view” that does little to initiate and support citizen engagement (Sieber and Johnson 2015). Bates (2012) echoes this research, stating that open government and open data initiatives, instead of redistributing power inequalities in relation to data, reinforces neoliberalist ideals and creates a free data-grab for those already in a position to use data. Likewise, literature also demonstrates how access to information is complicated by social, economic, and political issues, including digital inequalities (Castro 2014,
Norris 2001), policies related to information communication technologies (ICTs) (Zuiderwijk and Janssen 2013), and privacy (Garson 2006, Halstuk 1999, Pasquier and Villeneuve 2007).

Research into open data is relatively new, and within research specifically into municipal open data in Canada there has been little attention paid to scales. Rather, research that has discussed open data at different levels of government has largely taken scale as a natural organizing principle. Two recent evaluations of open data in Canada demonstrate this lack of attention to scale (Currie 2013, Roy 2014). In Currie’s (2013) evaluation of Canadian municipal open data, he makes little mention of scale although his research is very scalar. He states that the purpose of his inquiry is to abstract from municipal case studies to create a national narrative of municipal open data in Canada (2013, 71). When speaking of governance, Currie notes that “our usual understanding of governmental powers as existing on set scales (local, regional, and national) is problematic because this conception does not consider the multiple scales upon which nongovernmental actors can engage (Cox 1998)” (Currie 2013, 9). In this regard, Currie’s discussion of scale, which does not extend beyond this, maintains that scale is a set organizational model in which people can act in multiple ways and at multiple levels. In this regard, there is very little attention paid to the role of scale and how it functions as a way that people conceive of their world.

Similarly, Roy’s (2014) assessment of open data and open governance in Canada lacks attention paid to scales. Although he is comparing and contrasting open data and open-governance initiatives across different levels of government, little attention is paid to how scales function. Rather, they seem to be understood as natural organizing principles. Roy argues that they need to align the open data and open-government strategies and governance mechanisms at the “localized, subnational, and national” levels (Roy 2014, 414). Here, Roy attempts to differentiate between the hierarchical structure of governmental bodies (i.e., municipal, regional, provincial, and federal) and one that broadens the inclusion criteria for each scale category (i.e., local, subnational, and national), but instead he creates a functionally similar hierarchical scalar strategy. Differences and inconsistencies are noted at different levels, but there is no attention paid to how scales may be implicated in these differences.

METHODS

Interviews were conducted with ten participants from the following municipalities and districts1 of western Canada: cities of Victoria, Vancouver, Kelowna, Penticton, Vernon, and Edmonton, and regional districts of Central Okanagan and Okanagan-Similkameen (see Table 1). We chose the Okanagan as the primary area of study for smaller municipalities because the region demonstrated five open data initiatives that were evolving at different geographic and population scales as well as being at different stages of development. We chose the cities of Victoria, Vancouver, and Edmonton as examples of larger municipalities and because they are considered early adopters and leaders in Canadian open data, and are partners of the Canadian Geospatial and Open Data Research Partnership. We chose to include regional districts because they provide services to people that

<table>
<thead>
<tr>
<th>Municipality/District Participant</th>
<th>Population Size1</th>
<th>Ordinal Measure Based on Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>City of Edmonton</td>
<td>877,926</td>
<td>1</td>
</tr>
<tr>
<td>City of Kelowna</td>
<td>124,378</td>
<td>4</td>
</tr>
<tr>
<td>City of Penticton</td>
<td>33,964</td>
<td>8</td>
</tr>
<tr>
<td>City of Vancouver</td>
<td>648,608</td>
<td>2</td>
</tr>
<tr>
<td>City of Vernon</td>
<td>40,497</td>
<td>7</td>
</tr>
<tr>
<td>City of Victoria</td>
<td>84,793</td>
<td>5</td>
</tr>
<tr>
<td>Regional District of Central Okanagan</td>
<td>195,523</td>
<td>3</td>
</tr>
<tr>
<td>Regional District of Okanagan-Similkameen</td>
<td>81,841</td>
<td>6</td>
</tr>
</tbody>
</table>

1 In British Columbia, cities are defined as having populations of 5,000 or more; districts are defined as being larger than 800 hectares and having a density of less than five people per hectare; towns are defined as having populations between 2,500 to 5,000; villages are defined as having less than 2,500 people; regional districts have been created to service populations that are not included in other defined municipalities or districts and range in size and population.
are not included in municipal boundaries, and are custodians of significant datasets within British Columbia.

All participants were directly involved in their municipalities or district open data initiatives. Approximately one-hour interviews were conducted in person, except for those in the city of Edmonton, when the interviews were conducted using Google Hangouts, and the regional district of Okanagan-Similkameen, where they took place by telephone. Participants were asked to describe the open data initiative that they are involved in; the benefits and challenges that have occurred throughout the process of developing an open data initiative; how they see open data developing in the future; and how they feel they compare to other open data initiatives (see Table 2).

All government personnel interviewed were aware of the publicized potential benefits of open data as having positive economic, social, and policy impacts on governments, organizations, and citizens alike (Government of Canada 2014b).

In assessing government personnel’s perception of the efficacy of open data in their municipalities or districts, we understand scale to be a socially constructed organizational tool that naturalizes and sediments sets of sociospatial relationships through everyday practices (Kaiser and Nikiforova 2008). In this sense, we are looking at how government personnel are conceiving of the usefulness, benefits, and challenges of an open data initiative in relation to scale.

We discuss three issues to emerge from this research that represent the way that government open data initiatives shape the interaction between governments and citizens. These include not knowing who open data users are; the jurisdictional nature of the way that data is managed; and the way that governments conceive of what constitutes open data.

Table 2. Participant interview questions

<table>
<thead>
<tr>
<th>Questions:</th>
</tr>
</thead>
<tbody>
<tr>
<td>How is open data used in this particular governmental body?</td>
</tr>
<tr>
<td>What are the sorts of policies around open data in this governmental body?</td>
</tr>
<tr>
<td>What are the benefits and challenges of having an open data platform?</td>
</tr>
<tr>
<td>How is data shared within your government, with other governments, and with the public?</td>
</tr>
<tr>
<td>What are some differences or similarities between open data within this government and others? What about at different levels of government (i.e., municipal versus district versus provincial versus federal)?</td>
</tr>
<tr>
<td>Who are the users of open data? How do you know who users are?</td>
</tr>
<tr>
<td>What do you see as the usefulness of open data in your government?</td>
</tr>
<tr>
<td>How would you perceive the usefulness of open data from a user’s perspective?</td>
</tr>
<tr>
<td>Is there a particular way that open data is packaged?</td>
</tr>
<tr>
<td>Are there any limitations to the way that open data can be used? Accessed?</td>
</tr>
<tr>
<td>Do you think that the type of data offered will change over time? Do you think it will become more diverse or less?</td>
</tr>
<tr>
<td>In an ideal world, how would open data be used? How do you foresee its use in the future?</td>
</tr>
</tbody>
</table>

DISCUSSION

Open Data Users

The interviews highlighted the complaint that participant municipalities and districts are unaware of who actually uses the data and to what end-use it is being applied. As service providers, one of their goals is to provide worthwhile and cost-effective services that actually are being used. Because there is no direct way to track open data users, it is difficult to quantify the value of that open data. Similar to results in other cities, all participants said that governments have minimal evaluative tools running on their Web sites that gather information about what happens on the Web site (number of clicks, time spent on pages, and download counts). However, they have no ability to actually understand the use, as well as the impact, of this data. As the city of Penticton states, “there’s no way of getting feedback of what people are using it for.” Currently, the only way for a government to know how data is being used is if the user reports it. “Unless they tell us that they’ve used it, like on Twitter . . . we don’t know” (City of Victoria). In this sense, it remains difficult to gain a consistent idea of how open data is being used.

The open data users that are known to municipalities typically come from within the government or from businesses that leverage data for a product. Some examples that municipalities shared were ReCollect, Placespeak, and Bing Thom Architects. Open Data use by individuals in these cases is simply unknown. We do know that digital literacy plays a role in who has the requisite technical skills to manipulate open data in its raw form. In many cases, it is governments, businesses, and organizations that already have the hardware, software, and other resources required to utilize, repurpose, and analyze open data, as well as having access to people with the technical skills to manipulate the data.
The average citizen may not possess all the required resources and skills. However, that does not mean that citizens cannot benefit from the products of those labors, or that governments cannot develop and offer tools that make the manipulation of open data more accessible.

While there is no direct way that governments track open data users, some governments have made efforts to find out how open data is being used in other ways. The city of Victoria notes that developing relationships with local data groups and advocates was one way to help understand the needs of the community and how data is being used. Relationships additionally help governments make better decisions about their open data initiatives (such as learning which datasets should be prioritized for publication). In this case, feedback is important for governments to provide better services. Additionally, because of user feedback, the city of Vernon built an app for consultants that allowed consultants to access frequently used datasets. That not only saved time for government employees who had to fulfill information requests, but also provided better services to clients. A more novel approach has been taken by the city of Vancouver, which has implemented a crowdsourced approach to update the Cultural Spaces dataset in its open data catalog. Users add cultural spaces to the dataset, which then are confirmed by city staff. This approach to data management not only aids the city of Vancouver in gathering data, but it also allows citizens to define what cultural spaces mean to them.

Although there are few ways to track the users of open data, governments are able to glean some idea of users through data requests. In general, smaller municipalities report that the primary identifiable users of open data include local contractors, developers, and consultants. However, the cities of Victoria, Vancouver, and Edmonton have additional examples. Both Victoria and Vancouver mentioned hackathon events that they were involved with, noting that it was “a very useful and valuable exercise to be able to connect [with] businesses, data custodians, and with the developer as well, and anybody who’s interested in data” (City of Vancouver) and that it was a practical way to help different sides of the open data movement share expertise and collaboration.

For governments to unlock the full potential of open data, they need to be able to understand who the users of open data are. Without this knowledge, they are unable to effectively provide useful data to their users. One way to be better informed about user needs is to form relationships with open data communities. Secondly, as the nature of citizen-government relationships are changing, governments will need to change the way they communicate and develop a relationship with their citizens. Understanding who and how open data is being used does not have a clear answer, but will require new and novel approaches to citizen-government relationships.

Incorporating civic-engagement practices into government open data platforms also can greatly benefit open data initiatives. Currently, open data portals tend to have few mechanisms for users to engage with government about open data. At a minimum, users can send e-mails with feedback to the open data team; while at most, users can comment directly on datasets. In these cases, it is difficult to engage in conversation about open data with these limited means of communication. It seems that most conversations about open data currently are occurring outside the government sphere; by bringing that conversation to the individual data portals, or the source, valuable engagement with citizens, governments, and data can happen in a streamlined and relevant space.

Data Jurisdiction
Municipalities and districts recognize the benefits of being able access datasets from multiple government sources in a single repository. Jurisdiction of data, however, affects the types of services and data that a municipality or district can offer. We use the term jurisdictionality to describe how the authority of the collection, management, and publishing rights of data is divided by multiple governments into overlapping geographic areas.2 Because of data jurisdiction, it becomes more difficult for a governmental body to provide data from outside sources in its own open data portal because it does not own the rights to that data. For example, the regional district of Central Okanagan describes what a user has to go through to get data that covers the entire Central Okanagan:

One of the challenges is trying to put some of the pieces together when you have different municipalities providing data. For example, if you wanted to put together a full set of all of the cadastral lines, which basically are just your lot lines, for our area, because Kelowna and West Kelowna aren’t in our memorandum, basically they do their own data, I can’t provide that data online. . . . you would have to go to our site, download the Lake Country, download the Peachland and Westbank First Nations and ours, and then go to Kelowna and get theirs and then go to Kelowna and get theirs and then mosaic it together on your own.

Similarly, the city of Vancouver recognized the simplicity of being able to get all data from one place, but stated that if it started covering areas outside of Vancouver, then it would “have to make sure we have a reliable way of getting the data” from other municipalities. Not only is there worry that data availability might suffer, but also there are concerns about the quality of the data from outside sources. In this case, Vancouver could not guarantee that outside data met its standards. Ultimately, the biggest problem is Vancouver does not have enough capacity to manage its own data, let alone data outside its jurisdiction. In addition, because sites were starting to aggregate municipal datasets, there was no pressing need for cities to take on that responsibility. A similar problem exists in Victoria; with “13 separate municipalities in Greater Victoria,” it would be useful for citizens to be able to access data that crossed those boundaries. Yet the barriers to publishing data outside of one’s own jurisdiction prevent it.

Not only is data jurisdiction an issue of who houses the data,

---

2 For example, the process that determines who is the data authority for data that occurs at the coordinates 49.9054054-119.4360369, which is located in the city of Kelowna, the Regional District of Central Okanagan, and the Province of British Columbia.
Currently there are no operating standard practices in place to help governments provide data in ways that are compatible with other municipal, provincial, and federal open data systems. An open data standard is a set of best practices for publishing data in a way that makes it interoperable and compatible between organizations and systems. Bloom and Sieber explain that open data standards “facilitate the measurability, performance, and comparability of data through establishing common file formats, schema, and unique identifiers for data elements” (2016, 2).

In addition, there are no standard practices for publishing open data. Without a standard practice, municipalities and districts have to develop their own software for publishing open data, or outsource it to companies who specialize in open data portals. Either option can be a significant cost to a municipality or district, and it does not guarantee that it will be compatible with other open data portals. For example, the city of Victoria noted that it was only able to begin its open data initiative after it made serious upgrades to the city’s Web site several years ago. Similarly, the city of Kelowna stated that many of the advanced features that go beyond an open data catalog, such as built-in visualization and other GIS mapping features, are unavailable until it upgrades its internal servers to gain those functionalities. The cost and time to implement these newer systems can be limiting factors that can slow or even prevent open data initiatives from reaching desired goals. This is the case particularly in smaller and midsize cities in Canada. Because there is no standard practice for the software and hardware for open data portals, this can create barriers to the use of open data across multiple jurisdictions because each system has a specific way for managing its open data. Many governments have built their own GIS mapping systems and pull data from their existing GIS to populate their open data catalogs. In this type of setup, adding special functionality to its GIS often would require the city to build an app for it.

Several other options exist for open data portals. These alternative options are attractive because they typically do not require governments to upgrade their own hardware or software. Some municipalities are using a form of Environmental Systems Research Institute’s (Esri) Web-mapping applications. Depending on the version of the software, there is a certain amount of built-in functionality. However, the government does not have any control over the software and cannot add or modify functionality without vendor support, which often is very expensive. Another option is using CKAN, an open-source software that provides a substantial body of the code needed to establish an open data portal, but also allows governments to customize it to their needs. However, it requires technical expertise to deploy. Socrata is a prebuilt, vendor-supported platform that provides all the software and hardware needed. This option reduces the need for staff to maintain and host the data locally, for Socrata hosts and manages the data. This system, however, reduces governmental control over the back-end processes of the platform and licensing costs are associated with using the software.

In a time when civic open data standards have not uniformly been established, data jurisdiction creates challenges to using data from multiple sources. As Bloom and Sieber (2016) note, one of the greatest benefits to open data when considering standards is making it more interoperable. One of the major benefits of open data is that it can be used for any purpose; however, when that unrestricted creativity is hampered by data being collected, constructed, and configured in uncomplementary ways, open data loses an element of its accessibility and usability. Ultimately, interoperable open data is data that is greatly more usable than unstandardized data. When open data can easily be mashed up with other datasets, users save time and effort in converting it to a usable format. It becomes less of a matter of what standard to use, but more of choosing a standard, adopting it, and collectively developing that standard further through best practices.

Governmental Culture of Data

One of the biggest challenges that governments face internally is changing the traditional model around the culture of data. Within Canada, this traditional governance model focuses on collecting and keeping data primarily for internal use (Dawes 2010, Doberstein 2013). As previously discussed, although Canada’s Action Plan on Open Government promises to move away from this traditional model, so far it has continued to hold on to it in practice (Doberstein 2013). An open-government model, on the other hand, focuses on making data more available to citizens, encouraging governmental transparency and accountability (Sieber and Johnson 2015). Because an open-government culture is so new and different, respondents did not yet comprehend it, understand how to implement it, or did not see the value in it.

The city of Penticton gave an example of how it could provide data about snowplow allocation and deployment, but does not because of reservations about what the release of that data would mean for the city in terms of service provision expected by the public. It noted that there is fear, by some, that releasing data might lead to an overwhelming number of public requests for services. Ultimately, Penticton said, “there’s so much momentum [needed] to get anything changed.”

Although budgetary concerns over open data have been cited as a barrier (Janssen et al. 2012), the city of Victoria stated that such fears are secondary to understanding the value of open data. The interviewee noted, “I don’t think it’s a budget issue, so much as it is an understanding issue. Not understanding [open data], not getting the value, and, as a result, not prioritizing it during their day.” The city of Vancouver agreed and added that staff turnover also can have a big impact on the continued understanding of open data. In Vancouver, most of the original group of staff that began the open data initiative has retired. The city noted, “There were really good senior champions to help drive the program. It’s tough when you don’t have as much senior support and you’re still trying to make the case to show the benefit” (City of Vancouver).

The city of Edmonton says that open data is “a complete 180” for most people. In a traditional view, government has complete ownership over data, but open data methodology states that the information is public, and should always be public unless it is...
legally private. Even when there is council support and the presence of an open data policy, the city noted that it was a really “big challenge to get people to think differently about their data.” In addition, “it’s a culture and it’s going to take another generation or so to change the perception of who owns the data” (City of Edmonton).

**Limitation of Size**
For small municipalities, scale becomes a limiting factor in determining the efficacy of open data initiatives. Following Kaiser and Nikiforova on the social constructedness of scale, stating that scale is a powerful “way of seeing, thinking, and organizing knowledge about the world around us” (2008, 538), we find that municipal and district governments perceive that scale directly affects how effective open data can be to a municipality. In this sense, the effectiveness of open data can be understood as a quantity of usage: The greater the usage, the more effective the open data initiative. The potential or lack of potential for open data is understood through several variables, such as resources and capacity of the government, interest by citizens to use open data, and the needs of the community. Smaller municipalities and districts see larger municipalities as inherently possessing more of the necessary variables because of their increased scale. While it is true that the scale of a municipality can increase the capacity and resources of an open data initiative, we argue that in the context of scale, smaller-scale municipalities’ understanding of size goes beyond this rationale. Indeed, most research in Canada is conducted on large municipalities, with populations of more than 100,000 (Gill 2016). The expectation is that of a reverse economy of scale: What works in the big cities, surely will function in the small.

When smaller municipalities are asked about what they see as the difference between a smaller and a larger municipality in terms of open data, they describe larger municipalities as having “successful” open data initiatives that are continually producing deliverables to their citizens. The understanding here is that large municipalities are actively producing results from open data while smaller municipalities are not. The city of Kelowna explains that larger municipalities “want to have something to show for it. We just haven’t really put ourselves in that category, we’re a little too lean to do that.” In this sense, smaller municipalities do not feel that they can use open data to produce a product or service for their citizens because of their size.

A core component of open data is transparency; this, however, was seen by some municipalities as not being present at the municipal level. The city of Vernon noted that there was a “push for transparency at higher level governments like provincial, but it hasn’t filtered down at least to the city yet.” The mentality is that eventually some of the benefits, such as transparency, will reach the municipal level as open data practices “filter down.” Overall, for smaller municipalities, the scale effect results in the perception that action and change in government occur at higher levels (such as national) and then move down through successive scales.

When looking at instances of open data use, although larger municipalities such as Victoria, Vancouver, and Edmonton have more examples, those examples were largely generated through internal initiatives (such as Edmonton’s Citizen Dashboard) or were partnerships between the government and an organization to provide a service (such as ReCollect or Placespeak). While this is not undesirable and several very useful things have been created by governments from their own data, it speaks to the inability to comprehend how open data is being used by the public on any scale. When it comes to understanding the uses of open data, both small-scale and larger-scale municipalities have the same problems. In this sense, smaller municipalities are seeing only other governments’ use of open data; however, the scale of municipalities does not necessarily correspond with an increase in nongovernmental, autonomous open data use.

In smaller municipalities and districts in the Okanagan, there is also an attitude that the benefits of open data will be experienced differently in larger and smaller municipalities. The regional district of Okanagan-Similkameen illustrated this when a spokesperson said: “I’m not sure how much that applies in our area, it’s just a different scale or just a different environment where they have so many restaurants within a certain area and stuff like that. I don’t know how much that will directly relate to what we do in our areas, but I’m sure there’s other uses there that I’m not aware of.” Here, the belief is that the benefits and potential uses of open data are intrinsically linked to the scale of the municipality or district. Similarly, the city of Vernon talked about how open data has been credited with empowering citizens to create their own apps and tools and that those benefits are more common in larger cities because there is more interest and greater specialized knowledge there. Although it is true that the makeup of each governed area is different (i.e., economic, social, political, etc.), and therefore the needs and interests of communities vary, smaller municipalities are not excluded from the benefits of open data. Rather, it is the particular understanding of scale that shapes how open data can be of benefit. This understanding points to the “filtering down” view of scale in which the top-down direction of change and innovation results in the belief that open data is more suited and beneficial at a larger scale because there it can be properly utilized. Smaller municipalities think open data works well in meeting the practical needs of consultants and developers but other uses of open data, such as app creation and governmental transparency, resonate less. For smaller municipalities and districts, the possibilities for open data are directly tied to this scale effect. The issue with this kind of scale effect is that it can act as a limiting factor for open data initiatives because of the way that governments perceive open data changes the way that they adopt open data practices.

Although the limitation of size is a major factor for smaller municipalities’ understanding of open data, larger municipalities think about scale differently. They do not seem to be limited by it. When speaking of data peers, large municipalities tend to look farther away, or to provincial, federal, and/or internationally at state jurisdictions, rather than at smaller municipalities. Smaller municipalities tend to look regionally for their peers.
Districts and municipalities within the Okanagan tended to look at Kelowna or Kamloops as peers, while Kelowna tended to look at Kamloops, but also Vancouver and Nanaimo. Smaller municipalities did not look beyond British Columbia, and tended to see their peers more regionally than larger municipalities, which were more likely to look outside of the provinces or to other prominent cities in the United States. For some larger municipalities, scale plays less of a role in their open data initiatives than other factors. According to the city of Victoria, the key to a successful open data initiative is the presence of an open data leader. The city noted that some larger municipalities could be doing more because they have more resources and capacity, but they are not because they do not have an open data leader. It is worth considering whether the presence of an open data leader is the reason larger municipalities interviewed had such a different view of open data.

POLICIES: ESTABLISHING OPEN DATA

The way that open policies are developed also reflects a municipality’s or district’s understanding of open data. Open Data policies are a formal way to cement open data initiatives and help reduce barriers and resistance to changes in data management by making open data initiatives part of a lawful process. Currently, it is uncommon in western Canada to have an open data policy. However, two governments interviewed—the city of Edmonton and the regional district of Okanagan-Similkameen—have added formal policies to assist and solidify their open data initiatives. In general, these policies state that data that is not protected under privacy legislation should be made public. Furthermore, they state that because the public, through taxes, has funded the collection of this data, they should have access to it free of charge. Several governments, in their open data initiatives, are working toward having a policy. For example, both cities of Vernon and Penticton are planning to implement an open data policy; Vernon already has an open data catalog, while Penticton still is working toward its first iteration of an open data platform.

Not all governments, however, are pro-policy. The city of Vancouver has chosen to not implement a policy at this time. Instead, through support from the council and education on the benefits of open data, the city hopes to encourage a more bottom-up approach to open data support. In this sense, the municipality is focusing its energy on changing the way internal staff understand data management, and is hoping to move employees toward understanding the importance of open data through that approach rather than through enacting mandatory policy. However, the council has passed a motion called Open 3:

[W]hat that basically says is that there are three components in there, open data, open source, and open standard. And for the open data program we focus mainly on open data, obviously. And what it says in the council motion is that if there’s no security, no privacy or confidentiality concern then basically we should release it to the widest audience possible and make it accessible. (City of Vancouver)

For some, the focus is on changing the internal culture around data so that a policy is not necessary, while others see instituting the policy first that makes open data publication mandatory, and then changing attitudes about data.

When governments enact open data policies, they often are embedded within broader policies on open government. Among the group of participants interviewed, one municipality and one district had specific policies relating to open data—the city of Edmonton and the regional district of Okanagan-Similkameen. Other governments, such as the cities of Vancouver and Victoria, had council motions supporting open-government and open data initiatives.

The city of Edmonton established its Open City Policy (C581) on April 14, 2015. This Open City Policy defines open government as a “philosophy of government which emphasizes the value of greater transparency and accountability, increased citizen engagement, and innovation and economic opportunities driven by these values and by the release of data and information” (City of Edmonton 2015, 1). In addition, the Open City Policy develops open-government principles from a wider civic perspective. This distinction between the two terms is important to understanding the overall policy.

The Open City Policy covers six topics. The first recognizes “information and data as a strategic resource” (2015, 1), meaning that both will be managed in a way that makes them easily accessible, because they are valuable decision-making tools. The second states that the city will make nonprivate data open by default, offering it in a variety of formats and media. The third topic states that the city will look for opportunities to create a more collaborative environment with citizens and to better engage them with the “design, development, and delivery of public programs, services, and policies” (2015, 2). Fourth, the city will use new technologies and innovation to deliver programs and services in multiple ways, including online and in person. Fifth, the city will actively remove accessibility barriers and create new ways to have collaborative relationships between the city and its citizens. Sixth, the city will “work with other public and private-sector organizations for the advancement of Open City principles” (2015, 2). The policy for the regional district of Okanagan-Similkameen (RDOS) is not an open-government policy; rather, it is a policy on the release of GIS data. The Enterprise Unit Data and Services Policy was enacted on December 10, 2009. In this policy, references to the release of data do not name it “open data” specifically. Instead the policy notes that “RDOS will provide select digital data to the public, government agencies, or contractors at no charge” (2009, 1) if there is no conflict with the Freedom of Information and Protection of Privacy Act. Alternately, if the data is from a third party, the regional district will only release it with permission from that party.

At this time, policies for open data are uncommon; only two policies were identified by interview participants among the districts and municipalities. These serve as examples of solidifying open data into law at the district and municipal levels. For the city of Edmonton, open data has been placed under a larger
open-government policy that aims to increase transparency, accountability, civic engagement, innovation, and economic opportunities. Alternatively, the policies of the regional district of Okanagan-Similkameen are strictly related to geospatial data with a main focus on the requirement that data be released free of charge. These serve to highlight the differing approaches to open data. In this sense, the policies that have been put into place are reflections of how governments conceive of open data and thus shape how open data initiatives are carried out. For the city of Edmonton, open data is being used as a tool to help deliver particular open-government goals, while the regional district of Okanagan-Similkameen uses data primarily as a way to reduce workloads for government employees.

CONCLUSION

Governments, as the stewards of large amounts of data on the public they serve, are in a time of transition. With the increasing adoption of open-government principles in recent years, the release of freely available public data continues to grow. Governments must redefine how they collect, manage, and publish data. As governments begin to release data to the public, they not only change the way that they handle data, but also how they interact with citizens. Traditional methods of understanding citizen engagement do not work well with identifying open data users. This creates a situation where governments have very little knowledge of external and internal end-user data use and, as a result, have difficulty making informed decisions about their open data initiatives. New ways need to be created to understand how data is being used. Some examples can include civic-engagement processes through creating relationships with open data users, and creating spaces, both online and offline, where conversations about open data can occur. Secondly, governments may have to redefine what constitutes a successful open data initiative because traditional measures of “success,” for example having deliverables, may not adequately measure the potential usefulness of open data, or the engagement of citizens, organizations, and businesses with that data.

Additionally, the jurisdiction of data into small areas of control that clearly reflect and replicate different governing bodies has created situations in which data often is not interoperable with data outside of that particular jurisdiction. Governments need to come together to find solutions to increasing the interoperability of data as well as the accessibility of data. One solution is to create standardized practices for the collection and publishing of open data, which can increase the usability of publicly available data.

The way that governments understand and interact with open data shapes how initiatives develop. Internally, there currently appears to be some friction when municipalities and districts adopt new ways of collecting, managing, and distributing data. Our research shows that this struggle is present at all sizes of municipal government whenever changes occur and represents serious challenges to open data initiatives. The culture around data ownership is in a transitional period between traditional data-management models and newer open-government models that include open data methodologies. Outreach is needed for governmental bodies to understand the benefits of releasing their data freely to the public that includes tangible examples of open data benefits. This presents a challenge to governments and organizations in developing strategies and tools for evaluating the success of open data.

In addition, there are fundamental differences in the ways that governments conceive of data at different scales of government. Smaller municipalities and districts tended to articulate that the perceived benefits of open data do not entirely apply to them because they were too small, and thus structured their open data initiatives accordingly. For example, some participants interviewed felt that their municipalities or districts were too small to have any app development take place, or that governmental transparency was something that could occur because of their open data initiatives. Larger governments did not share this perspective and instead felt that open data could, in part, accomplish goals of transparency, accountability, civic engagement, and economic and social innovations. These particular understandings at the intersection of scale and open data affect the way that open data initiatives have developed in these municipalities and districts, as well as their plans for future development. Additionally, these different understandings of open data also are apparent in the policies relating to open data. On one hand, the city of Edmonton’s Open City Policy focuses on broad open-government principles, while the regional district of Okanagan-Similkameen’s policy focused exclusively on the free release of geospatial data to the public.

Each government is at a different stage of engagement with its open data initiatives and policies. Some have only superficially developed their open data platforms, while others have been developing their open data infrastructures and datasets for years. We feel that the current use and distribution of open data remains in its infancy, while the future holds enormous potential; however, for this to occur there is a clear need for municipalities of every scale to more wholeheartedly embrace the changes required to realize this potential. By understanding the perspectives of government personnel in charge of open data, we can better understand the trajectory of these initiatives. As citizens, we then can help governments develop open data in a way that benefits more of their users and more greatly reflects the needs of their citizens.

Acknowledgments

The authors would like to acknowledge and thank members of the municipalities that took part in this research, as well as the project funder, the Social Sciences and Humanities Research Council of Canada Geothink Project: SSHRC 895-2012-1023.
About the Authors

Mark Gill has a Master of Arts in interdisciplinary studies from the University of British Columbia and was a graduate researcher with Geothink. His research fields include open data, social justice, and community engagement.

Community Culture and Global Studies
University of British Columbia Okanagan
mark.richard.gill@gmail.com

Jon Corbett is an associate professor in community, culture, and global studies at UBC Okanagan and the director of the Institute for Community Engaged Research. His community-based research investigates cartographic processes and tools that are used by local communities to help express their relationship to, and knowledge of, their land and resources.

Community Culture and Global Studies
University of British Columbia Okanagan
jon.corbett@ubc.ca

Renée Sieber is an associate professor in geography and environment (jointly appointed) at McGill University. She conducts research at the intersection of social theory and software architecture. She currently leads Geothink, a cross-Canada project of 26 researchers and 30 partners, which investigates how new technologies and data are shaping city-citizen interactions.

McGill University
Environment McGill University
renee.sieber@mcgill.ca

References


City of Edmonton. 2015. C581 open city policy. Corporate Services Department.


Currie, L. J. 2013. The role of Canadian municipal open data initiatives: A multicity evaluation (Master’s thesis). Queen’s University, Kingston.


(Footnotes)


Open Data Community Maturity: Libraries as Civic Infomediaries

Pamela Robinson and Lisa Ward Mather

Abstract: Since the time of ancient Greece, the public library has been an important civic institution in support of democracy. With their provision of free, accessible information at the community scale, public libraries have a long history that tightly aligns with the goals of progressive local governments now engaging in open data efforts. While libraries and librarians have long positioned themselves as leaders in animating open-information provision, local government open data efforts, thus far, have been more focused on how city hall can make data available rather than engaging people with its use. This paper considers the manner in which public libraries are well positioned to act as intermediaries that render this data useful to the general public. The authors employ the term civic infomediaries to describe how these nonpartisan public institutions have the ability to connect community members with open data in a meaningful and civically engaged manner. This paper first explains the importance of ensuring that open data is accessible and useful to a nonexpert public and then explores the ways in which libraries have the capacity to connect users with open data resources, provide access to necessary technology, and act as a hub for civic digital activities. Ultimately, this paper argues that local government staff with open data responsibilities could be more active in engaging in partnerships with public libraries to create a more dynamic and robust open-government ecosystem.

INTRODUCTION

The North American open data movement has seen enormous growth over the past five years. While many early open data policies focused “primarily on making data available” and were first perceived as a fringe idea in the minds of technology advocates, governments worldwide have now declared themselves “open by default” (Zuiderwijk and Janssen 2014, 27). The new norm in government is to make data not only readily available, but accessible, standardized, and free of charge (Janssen et al. 2012, Code for America n.d.). The process of mainstreaming or normalizing open data has triggered a number of important outcomes. One such outcome is the emergence of a new way of thinking about open data efforts as “open data ecosystems” and, more broadly, “open-government ecosystems” (Pollock 2011, Harrison et al. 2012). According to the ecosystem understanding of open data, making data and information open are first steps in the creation of a connected web that leads toward more transparent, accountable, open, and democratic governance. One effect of the growth of the open data ecosystem and the corresponding growth in the availability of open datasets is the rise of the civic technology movement.

Civic technology involves the use of open data, among other datasets, as the key information input into mobile-app technology development with a civic focus (Johnson and Robinson 2014). These applications have a variety of goals, such as helping people gain better access to municipal service delivery (e.g., transit trackers such as NextBus), give input into and receive information about municipal government activities (e.g., Textizen, an SMS public-engagement service), and learn what public facilities are close by (e.g., Seattle Park Finder). Together the ecosystem-thinking approach and the civic technology sector offer new engagement techniques and insights into how to connect people with open data so that its value can be unlocked. However, they also serve as reminders that the provision of open data on its own will not necessarily lead to desired change. While entrepreneurs experiment with ways to reap economic value from open data, it is unclear who will meet the challenge of putting open data to work on behalf of the public good.

In the early days of the Internet, the term infomediary was coined to describe people who “leverage the Internet to unite buyers and suppliers in a virtual marketplace to facilitate a transaction” (as cited in Anttiroiko 2008, 3,674; see also Janssen and Zuiderwijk 2014). This market focus, however, does not sufficiently capture the civic or public-good focus of the open data, open-government, and civic technology efforts that we are witnessing today. As open data and open-government ecosystems evolve and mature, deliberate efforts are being made to help unlock the civic value of open data.

This paper argues that a new kind of infomediary—the “civic infomediary”—is needed in the open data ecosystem, and that municipal government open data efforts should be deliberate about partnering with public libraries whose efforts in this direction predate our current digital age. A civic infomediary is a person or organization that connects community members with open data so that public value can be derived from the data. The shift from open data provision (e.g., catalogs) to open data ecosystems necessitates a reconsideration of the roles played by different actors in open data efforts. The civic infomediary is an important actor in open data and open-government ecosystems that serve to engage ordinary people, and not just experts, in open data use. This paper begins by discussing the importance and difficulty of making open data accessible to a wide range of
people. It then turns to the efforts of one type of public institution—the public library—and positions libraries’ work as that of a civic infomediary. Libraries are nonpartisan public institutions that offer physical places for people to access learning (Jaeger and Bertot 2011). Libraries, which have long functioned as centers for civic engagement and have fueled democratic movements (Kranich 2010), are increasingly reimagining their future beyond books, bricks, and mortar. In recent years, many libraries have become sites for makers and hackers to manipulate technology and data. Interestingly, library staff, scholars, and researchers see libraries’ growing involvement with digital tools and information as a continuation of their long-standing work in technology-service provision and civic engagement. For instance, in 2003, Lowry found that libraries were adapting well to the “emerging world of networked information” and that students continued to depend on the library, albeit in “significantly different ways” (ix). However, this conception of libraries’ role—as centers for civic engagement and as digital-information hubs—is not necessarily recognized outside nonlibrary communities, including the communities of civic technologists and municipal open data staff and researchers. This paper, which identifies public libraries as important but underutilized potential partners in the open data and open-government ecosystems, is likely to be of most interest to these nonlibrary communities. This paper concludes with recommendations for future research and reflections about the role(s) that civic infomediaries might play in working with government staff whose work involves urban and regionally specific geospatial information.

**ACCESSIBLE OPEN DATA**

With governments embracing the “open by default” approach to open data, there is a migration from open data as a novel concept to a practice with more endurance and permanence. However, government staff and open data proponents and users are increasingly recognizing that making data open is only the beginning; the true value of open data comes from its deployment (Johnson and Robinson 2014, Sieber and Johnson 2015). The open data movement currently is shifting focus from pushing for data to simply be open, to calling for organizing and structuring the data in a usable, accessible, and in-real-time format (Sieber and Johnson 2015). In the community of technologically proficient open data users, these datasets are finding their way into for-profit and not-for-profit apps (e.g., Code for America). In communities with well-established open data programs, there also are new and more frequent efforts to make even more open data accessible and usable to nonexpert users (Bartenberger and Grubmüller-Régent 2014).

Early research on open data identifies a wide range of barriers to its use, including institutional challenges, the complexity of working with data, capacity barriers in all user groups including government staff and in the general public, information quality, and legislative barriers (Janssen et al. 2012, Zuiderwijk et al. 2012). Research shows that the general public has a hard time working with open data because it is in a format that is inaccessible to anyone but experts, or because members of the public do not have sufficient training in digital tools, or both (e.g., Currie 2013). David Eaves (2014), a longtime Canadian open-government advocate, is consistent in his message that open data needs to be organized and shared in a way that is easily searchable by a wide range of people, not just specialists. Working with open data requires a certain degree of computer and statistical literacy (Magalhaes et al. 2013). These barriers to access and use mean that, as Jacknis (2014) notes, “open government, particularly the open data initiatives, . . . is not always very useful to the average citizen.”

It might be tempting to step back and suggest that open data really does not need to be usable or accessible to all and that it would be more efficient to focus open data efforts on facilitating expert-user access and needs. This temptation is problematic in the civic context for two reasons. First, to leave open data in the realm of the expert undermines its transformative potential. The prioritization of expert use truncates the possibilities of civilian ingenuity that comes when an intrepid ordinary person mashes up data or creates an app that responds to public needs. Second, there is a reciprocity between open data and open government that needs to be explicitly recognized. Open data is widely recognized as having the potential to support transparency, accountability, and accessibility (the pillars of open government) and governments that are open have active open data efforts.

Open data is an input into open government. By manipulating these data, the public can better see and evaluate the work of government, which, in turn, allows them to advocate for or implement policy and service changes based on evidence from the data. Various organizations and advocates have argued that open-government services and data need to be “driven by citizen demand” (IFLA 2013, 18). Institute of Museum and Library Services Director Susan H. Hildreth argues, “We won’t have open government, accountability, or citizen engagement unless our citizens can find and use government data and records” (University at Albany 2015). When open data is deployed for all users, it can reinforce or buttress the open nature of government. It also is essential that governments ensure effective public involvement with open data lest they create a new barrier between citizens and government information (Gurstein 2011). If open data is released in a way that only facilitates usage by the expert community, then open government is only available for elites and the opportunity to make government open for all is lost (Gurstein 2011).

By tailoring data release to expert users, government also misses opportunities to make open data as useful as possible for both experts and ordinary citizens. There has been a growing awareness in the open data community that, thus far, there has been brilliant progress on the supply side—opening up data and multiplying tools and apps of all kinds. But there has been far less progress on the demand and use side. The result is that thousands of promising datasets, apps, and sites remain unused; and a great deal of creativity and energy has gone to waste. (Mulgan 2013)
Daniel Kaufmann of Revenue Watch uses the term zombie data to refer to the “enormous collections of data that lack purpose and insight” (Thomson Reuters Foundation 2013).

A related issue is the value proposition of open data as a driver of better access. Early research on civic hackathons revealed that some municipal government staff in Canada are being asked to demonstrate to their own organizations the value of making data open (Robinson and Johnson 2016). In these local governments, where a business case for open data is sometimes required to facilitate opening up more data, the social and economic value of open data only can be realized when there is a diversity of users (see Martin 2014). NASA underlines the significance of this connection between data and its users, stating, “the value is derived from consumer use of the data. There is no inherent value in idle data. The objective of this endpoint is to unlock the significant public investment in earth observation data” (Thomas n.d.).

There is a subtle shift here in the discussion about open data in which we move from “open by default,” with its implicit end goal of taking the data and making it available, to “open in play,” in which the data is open, findable, usable, and actually in use. As Sieber and Johnson (2015) note, when governments publish data (e.g., “data over the wall”), it is the first step in creating a broader open data ecosystem. The bigger goal is what Sieber and Johnson (2015, 311) frame as “participatory open data,” which occurs when governments and citizens engage in the coproduction of data. This active open data scenario has a greater social, political, and economic value, but it requires ongoing attention and effort. Speaking about civic hackathons as a means of connecting the public with open data, Harvey Low (2014), Manager of the Social Research Unit at the City of Toronto, reports:

So you need to begin thinking about it not being just about hackathons. While I think hackathons are great, you asked whether we are serving the needs of Canadians. What we’ve heard in the City is that they need an intermediary. They need an intermediary to take that data, such as that from municipalities or the province, and analyze it on their behalf.

When thinking about open data and open-government ecosystems and civic technology tools in a practical way, a question arises, which is: Who is doing this work? Or, in the language of political scientists: Who are the actors and what do they do? The role(s) of the people who work with civic open data are important because they remind us that these data are not transformative in and of themselves.

This recognition that data needs to be stewarded is not new. Hagel and Rapert (1997) are credited as the first to use the term infomediary during the early days of the Internet. They predicted that “companies we call infomediaries will seize the opportunity to act as custodians, agents, and brokers of customer information, marketing it to businesses on consumers’ behalf while protecting their privacy at the same time” (Hagel and Rapert 1997, 54). In their discussion, Grover and Teng (1997) argue that “by definition, infomediaries leverage the Internet to unite buyers and suppliers in a virtual marketplace to facilitate a transaction” (as cited in Anttiroiko 2008, 3,674). These descriptions call to mind contemporary companies such as Facebook and Google. The definition of infomediary has shifted slightly since the late 1990s. A more contemporary definition is: “an Internet company that gathers and links information on particular subjects on behalf of commercial organizations and their potential customers. Infomediary is a combination of info(rmation) and (inter)mediary” (Entrepreneurial Insights n.d.). However, it is unclear how these third-party, private-sector, customer-focused infomediaries help deliver public good in an open data ecosystem.

Infomediaries, as imagined previously, are paid third-party organizations working between governments and their citizens. While it is easy to imagine how these services could help accelerate open data use and app development, it is more difficult to imagine that community members will have universal and equitable access to data if they are required to pay for the services of an infomediary. This arm’s-length paid relationship means that the public is still not directly engaging with the data and that open data still is being used in, or mediated by, an expert system. When the data work is outsourced to an infomediary, the public misses the chance to roll up their sleeves and get their hands dirty with the data. This arm’s-length relationship usurps deepening public open data literacy and thus limits the potential for open data to lead to democratic change. If the goal of open data programs is to develop an active environment in which the public works directly with data to achieve social, political, and economic value, then these programs should plan to involve a new kind of infomediary—a civic infomediary—that supports direct public engagement with open data (see Table 1).

**LIBRARIES AS CIVIC INFOMEDIARIES**

Key challenges currently facing the open data movement relate to maintaining and organizing open data, and making those data accessible to a wide range of users. These challenges could be effectively addressed by an institution whose purpose is to

<table>
<thead>
<tr>
<th>Infomediary</th>
<th>Civic Infomediary</th>
</tr>
</thead>
<tbody>
<tr>
<td>By definition, infomediaries leverage the Internet to unite buyers and suppliers in a virtual marketplace to facilitate a transaction (as cited in Anttiroiko 2008, 3,674).</td>
<td>Civic infomediaries help make information and data accessible and understandable so that citizens/the public can make better use of it in pursuit of the public good/democracy.</td>
</tr>
</tbody>
</table>
steward information and connect communities with that information—the public library. Libraries have undergone remarkable changes as their users’ needs have shifted. The public library as “information hub, steward, and intermediary, has long been an important community resource [but] . . . [in the 21st century, the public library is increasingly being called on to take a more significant role in responding to community needs” (University at Albany 2015).

Recent research has begun to explore the changing role of public libraries. The Knight Foundation,1 the Information Policy and Access Center,2 the Aspen Institute,3 and the University at Albany’s Center for Technology in Government (CTG)4 all have initiated projects that shed light on how libraries can and should change as civic institutions. From this and other work on libraries, we can discern several key features that suit the library for the role of civic infomediary in the open data movement. The first of these is the library’s increasing focus on technology.

Although libraries are rarely thought of as places on the cutting edge of new technology, in fact, the Pew Research Center’s Internet and American Life Project concludes that many libraries are “transforming themselves into technology hubs” (Peterson 2013). Libraries have increasingly been adopting a range of roles in their communities related to digital technologies, such as providing computer access and digital services, including digital-literacy instruction. A particularly important role that libraries currently play is delivering e-government services (Jaeger et al. 2014). “Bookless” libraries have even been established, such as BiblioTech, the all-digital public library in Texas, and the Do Space library in Omaha, Nebraska. Do Space describes itself as a “technology library, a high-tech workshop, and an innovation playground for our community. It doesn’t matter how savvy you are or how much you know. At Do Space what counts is your desire to learn, create, explore, and invent. Our mission is to empower Do Space members through technology access and education” (Do Space n.d.).

As many libraries have begun to offer technological equipment and training, they also have become more “collaborative and interactive” (Coyle 2013). For instance, many libraries have hosted hackathons and established makerspaces, which are “workspaces that provide technological tools and are designed to facilitate collaborative work” (Fallows 2016). A 2013 web-based study found that 41 percent of responding librarians currently provide makerspaces or offer maker activities in their libraries and 36 percent of respondents planned to start makerspaces in the near future (Price 2013).5 Miguel Figueroa, director of the Center for the Future of Libraries at the American Library Association, explains the popularity of these initiatives, saying “makerspaces are part of libraries’ expanded mission to be places where people can not only consume knowledge, but create new knowledge” (Fallows 2016). However, this optimistic claim has been challenged by critics (Vossoughi et al. 2016, 208), who point out that “the mainstream discourse of making is also distinctly economic. Practices such as taking things apart, building new designs, and testing out solutions are valued insofar as they contribute to new technological and commercial innovations.”

Libraries also are well positioned to serve as civic infomediaries for open data initiatives because they are “open, public, neutral civic space[s]” that have a strong connection to their local communities (Hill and Mattern 2014). Libraries provide the physical infrastructure that can support their communities’ “informational and social infrastructures” (Mattern 2014). This support can come in a variety of forms, such as hosting events, providing meeting spaces, and allowing free and open access to a range of technologies. The Toronto Public Library, for instance, says that its local library branches “design programs and services to meet the needs of their local communities. For each branch, collections are carefully selected to meet local needs, custom programs are scheduled, and services are provided to ensure culture and language needs are met” (Martin Prosperity Institute 2013, 9). The Edmonton Public Library has been actively hosting hackathons and other data-related events, and its Open Data Mission Statement explicitly defines its responsibility in “working to increase open data literacy and encouraging citizens to unlock the value of open data” (Edmonton Public Library n.d.). This trend is not limited to larger cities: In 2017, Saint John Free Public Library, in partnership with the city of Saint John, New Brunswick, Canada—a community of approximately 67,000—hosted a series of four Open Data Book Club workshops on topics such as health care, transportation, and social planning, all under the theme, “Finding the Livable City” (Saint John Free Public Library n.d.). Because libraries are connected in this way to their communities, they are ideally situated to become hubs for civic engagement with open data. Many libraries already provide data-literacy instruction and Internet/computer access. These functions could be usefully expanded to include further access to and instruction in the use of open datasets to help communities answer the questions that are important to them. In 2016, for example, the Toronto Public Library organized an open data hackathon in which participants were invited to tackle several problems faced by organizations involved in the city of Toronto’s Poverty Reduction Strategy (Toronto Public Library, n.d.).

The last of the public library’s strengths from the perspective of the open data movement is its expertise in maintaining, organizing, preserving, and curating information, as well as teaching people how to use and understand that information. This expertise means that libraries are well positioned to help municipal governments solve the problems surrounding the organization, searchability, and accessibility of open data. Data librarian Celia Emmelhainz (2015) is excited about the potential of open data, but recognizes that the public will require help if this data is to become “truly accessible” to them. She argues that data librarians can introduce community members to open data, “opening up the world of data to the community, helping people to access public data, or hosting workshops on data skills. Here I’m thinking of things like scraping real estate data and visualizing it using infographic tools like impact.io. People don’t need a data genius as much as a data guide—and that’s what librarians are there for” (Emmelhainz 2015). Providing instruction in open data
literacy is especially germane if broader participation is sought from constituencies traditionally excluded from digital-knowledge production (i.e., non-technologically inclined individuals) with the aim of transforming urban open data into meaningful place-based knowledge (Mattern 2017).

One city that has been exploring a new role for the public library in its open data initiative is the city of Boston. Boston’s award-winning project entry in the Knight Foundation Library Challenge proposes “connecting the city’s open data program with [Boston] libraries” (Wood 2015). Boston’s Chief Information Officer explains that we have a massive volume of data that we release to the public through this web site, but it’s not very well organized and it’s not very useful. . . . Instead what we want to do is bring the skills of librarians, cataloging, curating, relating datasets to bear on this incredible information resource and turn it into something that can be truly useful as knowledge for our residents, for businesses, for government employees, and researchers. (Wood 2015)

In cities such as Boston, libraries could become the principal civic infomediaries between the data, the public, and government.

The example of Chattanooga, Tennessee, also provides some insight into how libraries can support municipal open data efforts. According to Nate Hill, the former Deputy Director of the Chattanooga Public Library, this library has become “both the open data platform for the city and the home field for civic engagement related to open data and civic geeketry in general” (Hill 2014a). The Chattanooga Public Library has taken on a central role in the city’s open data movement for several important reasons. For one thing, the library had available unused space—a fourth floor that was not being used. Another important factor is the city’s fiber-optical network, which gives people access to 1 Gbps Internet speeds at low cost (Hill 2014a). This network allows the library to provide the “infrastructure for [digital] knowledge exchange” (Hill 2014b).

Chattanooga’s political culture also contributed to the library’s ability to advance open data initiatives in the city. Hill (2014a) calls Chattanooga’s political climate “supportive and ripe for change.” Nonprofit and citizen groups also were crucial in Chattanooga, which is a Code for America fellowship city. Collaborating with the Open Chattanooga Brigade and the city of Chattanooga, the library now provides “open-government and locally relevant datasets as library collections” (Hill 2014b). The transformed Chattanooga Public Library is “a hub for civic innovation, a public makerspace, a Giglab, and perhaps most important a flexible beta space simply for trying new ideas in public” (Hill 2014a). The library has hired an open data specialist and continues to take an active role, along with its partners, in empowering “citizens with city data to tell stories, solve problems, and better interact with their government” (Open Chattanooga n.d.). Clearly, the Chattanooga Library has taken on the role of civic infomediary by providing support for community members who are interested in exploring and manipulating open data. As a result, it is helping to unlock greater civic value from the city’s data.

**CONCLUSIONS**

This paper has made abundant use of academic terminology such as open-government ecosystems and civic infomediaries. While the use of definitions and taxonomies is commonplace in the academic world, in a journal with a practitioner focus, it is the authors’ hope that this new labeling will be acceptable in workplace settings as well. For government-based geographic information staff, the distinction between infomediary and civic infomediary matters, particularly in a climate of austerity and tight budgets (Donald et al. 2014). In their recent research, Robinson and Johnson (2016) interviewed government staff who work with open data and found that a common challenge reported by staff was not enough time or resources available to both release old datasets and to continue the ongoing release of new open data. Also, in some local governments in Canada, municipal staff were being asked to make business cases to support government staff participation in events such as civic hackathons (Johnson and Robinson 2014). In other words, senior staff wanted to know what the return on investment of staff time would be and whether the hackathon, using civic open data, would result in the creation of a “killer app.” In governments with this business-case mindset, recognizing that there are important civic roles to be played by individuals and organizations to help bring open data “in play” confirms that a vital part of the value of open data transcends measurable financial deliverables.

Realizing the value of open data is an all-hands-on-deck proposition. Government open data has value inside and outside of government. Private-sector firms seeking to profit from open data can pay for the help they need, including the services of an infomediary. Governments have some internal data capacity, but civic infomediaries can help government achieve its open data goals. For example, Code for America serves as a civic infomediary when it convenes its brigade of coders to work inside government. Members of the public also could benefit from receiving this kind of support. Many nonexperts have a keen interest in open data and in fixing a problem but lack the data skills to get started. Here, civic infomediaries have an important role to play. This paper has focused on the public library to show—to open data community members—how its role is aligned with the animation and use of open data, but civic infomediaries can take many other forms. Moreover, this paper’s identification of the role and importance of civic infomediaries serves as a reminder that releasing data into an open data portal is only the beginning of a vibrant open data effort. As the role of civic infomediaries becomes better understood, urban and regionally specific geospatial information experts may themselves step in, or they may recognize the need for new partnerships and collaborations.
About the Authors

Pamela Robinson is the Associate Dean, Graduate Studies and Strategic Initiatives in the Faculty of Community Services, and an associate professor in the School of Urban and Regional Planning, Ryerson University. She is also a registered professional planner. As part of the geothink.ca research team, her research and practice focus on urban sustainability issues, particularly on cities and climate change and the use of open data and civic technology to support open-government transformations. She serves on the board of directors of the Metcalf Foundation. She is an editor of Urban Sustainability: Reconnecting Space and Place (University of Toronto Press, 2013) and Teaching as Scholarship: Preparing Students for Professional Practice in Community Services (WLU Press, 2016) and is a columnist for Spacing magazine (www.spacing.ca). She also serves on the Toronto Public Library's Innovation Council.

School of Urban and Regional Planning
Ryerson University
Toronto ON, Canada M5B 2K3
pamela.robinson@ryerson.ca
@pjplan

Lisa Ward Mather holds a Master’s of Urban Development from Ryerson University. She was awarded the 2014 Ryerson Gold Medal for the Faculty of Community Services, as well as the 2014 Canadian Institute of Planners' Student Award for Academic Excellence. She also holds a Ph.D. in English from the University of Alberta. Mather recently published an article entitled “Civic Crafting in Urban Planning Public Consultation: Exploring Minecraft’s Potential” in the International Journal of E-Planning Research (Volume 5, Issue 3, July–September 2016). She currently is working as a researcher on the Geothink.ca project, exploring the role of public institutions, such as public libraries, and civic technology in expanding the democratic impacts of open data.

1973 Stratford Ave
South Pasadena, CA 91030
lisawardmather@gmail.com

References


Gurstein, M. 2011. Open data: Empowering the unempowered or effective data use for everyone? First Monday 16(2-7): 1-5.


Notes

(Endnotes)

1 The Knight Foundation, motivated by a belief in the “centrality of libraries for building and maintaining an informed citizenry,” initiated a news challenge to explore the role of libraries in building more knowledgeable communities (Bracken 2014).
2 The Information Policy and Access Center (2011) has a libraries and e-government project that strives to determine “how best to bring together government agencies and public libraries. The project will also establish a set of best practices for the collaboration between government agencies and public libraries in the provision of e-government services to the public, which will enable their improved access and use.”
3 The Aspen Institute Dialogue on Public Libraries “provides access to an online community working together to address the transformation of public libraries in the digital age” (http://www.libraryvision.org/) and produced a report entitled Rising to the Challenge: Re-Envisioning Public Libraries. This report suggests that libraries can help people gain the skills and knowledge necessary to succeed today: “Economic, educational, civic and social opportunities are tied to a whole new set of knowledge and skills that barely existed a generation ago, and people without these skills or access to this information abundance are quickly left behind” (Garmer 2014, iv).
4 University at Albany’s Center for Technology in Government produced a report entitled Enabling Open Government for All: A Roadmap for Public Libraries, which “addresses growing questions about how ‘open government’ is influencing, and possibly transforming, the role of public libraries in their communities” (University at Albany 2015).
5 Respondents to the survey include “Librarians from 30 U.S. states, from Alabama to Wisconsin, . . . along with librarians from seven other countries (Australia, Canada, China, Denmark, Japan, the Netherlands, and the United Kingdom)” (Price 2013).
The Geospatial Contents of Municipal and Regional Open Data Catalogs in Canada

Edgar Baculi, Victoria Fast, and Claus Rinner

Abstract: Open data have become a symbol of government accountability and transparency, with data inventories growing rapidly in recent years. The open data phenomenon has been examined through communication, political, and legal lenses, but the geographic perspective, which is reflected in the amount of spatial content in the catalogs, has yet to be quantified and monitored. Analyzing the availability and accessibility of geographic information in municipal open data catalogs in Canada is an important prerequisite to understanding the potential for open data to inform decision making in the public and private sectors. Building on a previous survey completed in 2012, this paper investigates the status and trends of geographic information in Canadian open data catalogs by collecting an inventory of datasets and quantifying its growth. Municipal open data catalogs across Canada were indexed according to the type of their geospatial content. The growth of open data, and more specifically geospatial content, offers the public an opportunity to know, access, and engage with government infrastructure and services.

INTRODUCTION

The open data movement is a worldwide phenomenon by which everyday citizens are tapping into pools of data that governments are increasingly making public. The potential for transparency and public engagement has been the 32-calling card of the open data movement (Manyika et al. 2013, Sieber and Johnson 2015), giving way to transparency in public planning and decision making (Janssen, Charalabidis, and Zuiderwijk 2012; Scassa 2014). Through open data, everyday citizens and interested experts can review planning processes and form educated opinions that contribute to the decision-making process, what Linders (2012) described as a transformation of e-government to we-government. This bidirectional engagement could result in increased evidence-based decision making and integration of top-down and bottom-up approaches.

Given the novelty of open data, realizing its full potential is a work in progress. An important consideration in the open data movement is the supply side—who is providing open data, what data are being opened, and what support mechanisms are being used? On a global scale, several organizations are mediating the development of open data by recording the status and growth of open data content and initiatives, while providing standards and support mechanisms to institutions wanting to open their data. Open North, for example, is the steward of the Open Data Charter (ODC 2015), which offers strategic planning and tactical support for improving democracy through data provision. In addition, the Open Government Partnership has 70 (and counting) participating countries that have made a universal declaration to government openness and accountability (OGP 2016). In Canada, the commitment to open data was recognized through the establishment of the federal “Open Government” portal in 2014. As of the summer of 2016, the Canadian federal government had opened more than 185,000 datasets, and this inventory continues to grow at an unprecedented rate (Government of Canada 2016). With this high-level commitment to mediating open data, the open data movement is firmly rooted in national and international governance.

Supplementing this global perspective, locally, many organizations, meetups, and events are promoting the supply, use, and improvement of open data. A popular localized use of open data is hackathon events, in which government open data are used to develop software or applications that address a societal need, sometimes with the potential to create new jobs (Johnson and Robinson 2014, Lainchbury 2012). For example, the 2015 Canadian Open Data Experience invited participants to use Government of Canada open data to create new applications (CODE 2015). This was the largest hackathon in Canadian history with more than 1,300 participants in 340 teams from across Canada. It resulted in the creation of dozens of innovative apps, such as Career Path by Niew Labs, which help youth research and choose career paths based on detailed labor and demographic open data (Niew Labs 2016). The increasing interest in open data is supporting a similar increase in events that promote open dialogue between government workers, hackers, entrepreneurs, business professionals, and civil society. One unique example in Canada is the Open Data Book Club, where civic-minded citizens discuss, present, and use open data (Open Data Book Club 2016). Akin to a traditional book club, the group selects a monthly theme based on available open datasets. While the Open Data Book Club is maintained by Open Data Ottawa, there are several active clubs across Canada, including in Montreal, Toronto, and Prince Edward Island (Open Data Ottawa 2016).

With the importance of open data firmly established in all levels of government and nongovernment organization, it is necessary to shift focus to the condition and delivery of the data. Of particular interest is the GIS-readiness of data within existing open data catalogs. The aim of this paper is to explore the geospatial contents of municipal open data, building on the previous evaluation of open data by Currie (2013). This previous research provides a structure as well as benchmark statistics,
which allows us to track the growth of the Canadian municipal open data inventory. We present key open data concepts, which lay the foundation for understanding the GIS-readiness of these data. We also discuss the approach to surveying municipal open data catalogs. We examine the results highlighting the growth of geospatial content in the catalogs, and, finally, summarize our findings and provide an outlook on the future development of municipal open data in Canada and elsewhere.

RESEARCH CONTEXT

The Implementation of Open Data

Canada’s involvement in open data is supported by its membership in the Open Government Partnership, a global multistakeholder organization with the goal of government accountability and transparency worldwide, highlighted by the signing of the Open Data Charter by G8 leaders in 2013 (Scassa and Singh 2015). As early as 2003, national legislation in Mexico promoted government information to be proactively disclosed to the public online, which has resulted in a Transparency Portal with high numbers of visits since its first year of operating in 2007 (Davies and Lithwick 2010a, 2010b). In December of 2009, the United Kingdom put forward a report entitled “Putting the Frontline First: Smarter Government,” which contained commitments of public-data principles such as ensuring machine-readable formats, linked data forms, and unrestricted commercial use (TSO 2009). In 2009 and 2010, President Barack Obama declared government transparency and openness a goal, resulting in the launch of the data.gov portal, with federal agencies and departments being held accountable for providing high-quality data (Davies and Lithwick 2010b, Huijboom and Broek 2011).

Open data is based on the notion that government data should be free for all to use, including the permission for redistributions without copyright restrictions (Gonzalez-Zapata and Heeks 2015, Kassen 2013). “Linked data” refers to data that are published on the Web, are machine readable, and are linked to other datasets (Attard, Orlandi, Scerri, and Auer 2015). Governments collect data for the purpose of evaluation, justification, as well as developing services and programs (Sieber and Johnson 2015). Before data are processed, they are considered as objective, neutral, and without “meaning.” Through computer technology, data processing can help us understand complex problems in public and private uses. With open data, citizens have quicker access to the data underlying government reports and news reports, and gain the ability to create tools based on raw government data (Currie 2013). These technological and social changes have introduced new powers to the ordinary people, to better understand the inner workings of society (Bartenberger and Grumbmüller 2014, Manyika et al. 2013). Open data has roots in e-government, which aims to make government information, documents, and services accessible online (Linders 2012). Differences between the two approaches are based on engagement and enabling use, where the combination of open government and open data allows society to hold governments accountable when assessing costs and decision-making processes (Sieber and Johnson 2015). The open data movement advocates for the public values of e-government, to improve public confidence and trust in the government, through accountability and participation (Kassen 2013).

When it comes to understanding, standardizing, and effectively delivering open data, several views are put forward in structuring and packaging open data. Currie (2013) refers to Canadian open data activist David Eaves (Eaves 2009), who distinguishes three laws of open-government data:

1. If it can’t be spidered or indexed, it doesn’t exist.
2. If it isn’t available in open and machine-readable format, it can’t engage.
3. If a legal framework doesn’t allow it to be repurposed, it doesn’t empower.

Meanwhile, Veljković, Bogdanović-Dinić, and Stoimenov (2014) summarize the principles of the open-government idea:

1. Allowing citizens and entrepreneurs to access public-sector information and data in a uniform way.
2. Allowing the public to see government processes and operations.
3. Governments explain decisions made to their citizens, to take responsibility and show accountability.
4. Participation is increased by engaging citizens in the decision-making process.
5. Different levels of government, private institutions, and citizens can cooperate, improving collaboration.

Open-government data also can be understood from the four derived perspectives explored by Gonzalez-Zapata and Heeks (2015)—bureaucratic, technological, political, and economic. The bureaucratic perspective of open data addresses data policies, improving public service, and data management. The technological perspective deals with data infrastructure and technological innovation. The political perspective focuses on the right of access to public data, and its benefits for citizens and governance. Lastly, the economic perspective examines ways to generate profit from accessible data, mainly used by the private sector and entrepreneurs to create new products, jobs, and services (Gonzalez-Zapata 2015).

The Implementation of Geographic Open Data

Adding to the challenges of implementing open data is the addition or inclusion of geographic content within these data. Geographic data typically consist of elements that represent locations and properties of the environment and human-made objects and events (Currie 2013). Geographic data in raw form include geospatial coordinates, as well as facts, statistics, symbols, and characters (Judge and Scassa 2010). Integrity, and quality of any information that reaches the public through Crown works. With a view to facilitating the dissemination of government geographic data, GeoConnections (http://www.geoconnections.
org. These data often are the product of complex, expensive, and labor-intense processing. However, this does not make them original—a claim supported by the Supreme Court of Canada—meaning that these data have a place in open data initiatives (Judge and Scassa 2010). An entity of Natural Resources Canada called Geoconnections led the development of the Canadian Geospatial Data Infrastructure, consisting of open data licenses, standards, and services to improve the accessibility to geospatial data in Canada (Scassa and Singh 2015).

Identifying open geographic data is important as it is a step towards collecting and understanding the extent of geography within a catalog and within the greater open data movement. In an assessment of open data in the Netherlands, Welle Donker and van Loenen (2016) determine that geographical data outperformed nongeographical data. It can be argued that geography is one of the primary dimensions on which relationships between datasets can be established. As a subset of geographic data, we define GIS-ready data as those datasets that are in a machine-readable format, which can be opened in GIS. Citizens are aware of, and hold valuable information about, locations such as their home, work, or other familiar places. Thus, publishing geographic data as open data should be a priority for the public to integrate and interact with the data. Arguably, GIS-ready data close the gap between provider and user by minimizing the effort that the user needs to put into making the data useful.

One issue with the implementation of GIS-ready open data is the origin of datasets from various disparate government units, which can result in fragmented data storage (Conradie and Choenni 2014). Even when technical barriers of retrieving the data are overcome, the time and transaction costs associated with location, digitization, and release makes the data-publishing process lengthy (Conradie and Choenni 2014). Key recommendations by Zuiderwijk and Janssen (2014) to improve the implementation of open data include:

- Acknowledging the differences in data processing between departments,
- Asking community groups for input on which data to release, and,
- Allowing experienced departments to lead open data initiatives.

Government transparency includes the issue of making sure that electronic government data can be quickly located and retrieved, especially with the goal of capturing every relevant electronic government dataset (Jaeger and Bertot 2010). Making data available to the public as standard procedure may create a culture of transparency, innovation, and economic growth (Zuiderwijk and Janssen 2014). Open Data advocates summarize the benefits as increased transparency, innovation, collaboration, participation, and governance efficiency (Currie 2013).

Mediating Technologies for Geographic Open Data
Investigating the information technology associated with open data catalogs is necessary for the technology is what supports or limits what data can be found, scrutinized, accessed, processed, visualized, and, therefore, used (Zuiderwijk, Janssen, and Dwivedi 2015). Data not only need to be published but ought to be accessible to accomplish the activities of the end-users (Currie 2013). Several open data platforms provide a user interface for direct download of datasets in well-known formats (Sieber and Johnson 2015). Open Data portals add opportunities for independent developers to produce data-driven Web apps (Kassen 2013). The purpose of the data portal is to manage the metadata, data-format conversions, and visualizations (Attard et al. 2015). These open data portals become one-stop shops to maximize public-sector information, where published data can be explored and interacted with. According to Janssen and Zuiderwijk (2014), mediating between open data providers and users has led to a new infomediary business model, supporting the use of open data. A limitation with some data portals is the need to comply with local laws and bylaws that sometimes contradict their usefulness for the public (Attard et al. 2015). Data release often is challenged by copyright restrictions and privacy concerns (Conradie and Choenni 2014).

Digital information in the form of customized and interactive maps has allowed the public to experience their locations and neighborhoods in new ways, such as through geotagging of photographs (Currie 2013). Internet connectivity allows a wide range of users to make online maps in real time with minimal resources, allowing geospatial services to be extended to users regardless of time and location obstacles (Tsai, Choi, and Perry 2009). Web-based GIS enables users to create dynamic maps and reports, and create or edit data records from anywhere (Tsai et al. 2009). Evaluating open data technologies requires looking at search engines, Application Programming Interfaces (APIs), portal software, technologies for linking, assessing, analyzing, and visualizing the data, and storing the metadata (Zuiderwijk et al. 2015).

METHODOLOGY
Definition of GIS-ready Data Formats
Consistent with Currie’s (2013) study, we define geographic data as data in any format, such as Microsoft Office files (Word, Excel), PDF (portable document format), XML (extensible markup language), the content of which includes information that is pertinent to locations. These locations may be represented as street addresses, postal codes, coordinates, neighborhood identifiers, or other place names.

Furthermore, we define GIS-ready data as the subset of geographic data that are in machine-readable format and can be opened in current GIS software immediately, without further processing. For the purpose of this analysis, we counted only the Shapefile format, the KML format, and certain geospatial raster
Selection, Scans, and Records

This research was conducted as a Web survey with a subset of the criteria used by Crompvoets and Bregt (2003) in a survey of national spatial-data clearinghouses, such as the number of spatial datasets and the use of maps for searching. To scan and evaluate the geospatial content of municipal open data catalogs in Canada, we used Currie’s (2013) snapshot of 23 catalogs as of November of 2012 as a starting point. Our first scan was completed in December of 2013 and included nine of Currie’s (2013) cities, which at the time also were participating in the Geothink research project (http://www.geothink.ca). For a second comprehensive scan, we again prioritized comparability with Currie’s (2013) results. We also reviewed the list of Open Municipalities provided online by the government of Canada (Open Government 2016). On this basis, 46 cities and regional municipalities across Canada were identified for the January 2015 scan.

The scanning stages at both times involved the inventory of all datasets in each catalog and the available data formats. The scan of a dataset involved recording its name, description, data format(s), presence of map display, significant dates such as original publication and updates, as well as software platform used for the open data catalog. Some datasets had multiple formats. For example, a dataset on cemeteries from catalog (A) may include one GIS-ready Shapefile, while a similar dataset from catalog (B) of a different city may offer a Shapefile, a PDF listing, and an Excel spreadsheet. The scan takes this into account by including binary flags for every possible data format. Further examination through exporting and opening these files was conducted when needed to determine whether a dataset contained geographic data and GIS-ready data. A GIS-ready dataset also was automatically labeled as a geographic dataset.

The results are presented with reference to three groups of open data catalogs (see Figure 1 and Table 1). The first group are the 23 municipalities surveyed by Currie (2013) in November of 2012. Currie’s (2013) research took into consideration statistical datasets, economic datasets, along with geographic datasets. We compared Currie’s figures with the same 23 catalogs from a scan completed approximately two years later, in January of 2015.

The second group of catalogs is based on our preliminary scan in December of 2013, which included 11 cities. In this comparison, we are able to track the nine shared catalogs from the scans completed in November of 2012 (as reported by Currie 2013), December of 2013, and January of 2015, through their three-year annual growth. Note that Waterloo, Ontario, and Victoria, British Columbia, were not included in Currie’s study and therefore are excluded from this comparison.

The third group consists of all 46 catalogs from our most recent scan in January of 2015. This grouping provides the most recent and comprehensive snapshot of geospatial contents in municipal open data catalogs across Canada.

Platforms and Maps

The Web sites or portals in which the catalogs operate were themselves investigated. These software platforms may include map displays and GIS-like functions. Some catalogs offer maps in separate Web sites, which was noted in the inventory. The differences in platform determine how the user interacts with the open data. Nonspatial forms of interaction include viewing spreadsheets and creating infographics such as bar and line graphs.

Maps offer a visual aid to the user to better understand the location of the features in the selected datasets. Depending on the extent of GIS-like functionality provided by the platform of the catalog, users may be supported in discovering spatial patterns, especially when multiple datasets may be layered together. Map-related and GIS-related functionalities were important observations recorded for each municipal open data catalog.

FINDINGS

The Results of the Three Groups

Within the first group of 23 catalogs shared between Currie’s (2013) research in November of 2012 and our most recent scan in January of 2015, the grand total of datasets increased from 1,177 to 2,807 (see Figure 2). In other words, open data availability measured by the number of datasets more than doubled (+138%) over a period of approximately two years. With reference to geospatial content, 84% of the 2012 data were identified as geographic and 75% of those were categorized as GIS-ready. In 2015, 83% were geographic and 55% of those were GIS-ready. Interestingly, in both survey years, the proportion of datasets with geospatial content slightly exceeded the value of 80%, which is anecdotally referred to within the GIS community as the proportion of all data collected by government and businesses that have a geographic reference. The overall increase in published datasets from late 2012 to early 2015 in this group, however, came with a decline of the proportion of GIS-ready datasets, i.e., a slower growth of this type of dataset.

Among the second group of the nine catalogs shared between all three survey periods, the total number of datasets climbed
steadily from 573 (November of 2012) to 1,132 (December of 2013) and then 1,471 datasets (January of 2015) (see Figure 3). For the same time stamps, the proportion of geographic datasets stayed near constant at 79%, 77%, and 79%, while the proportion of GIS-ready datasets declined from 63% to 51% and then stayed at 52% of the geographic datasets.

The third group of catalogs consisted of 4,090 open datasets available in January of 2015 in 46 catalogs across Canada. Of these, 3,383 or 83% were geographic and, among those, 1,979 or 59% were in GIS-ready format (see Figure 4). The smallest proportion of geographic data in any of the 46 catalogs was 25% (one out of only four published datasets), followed by 28% (17 out of 61) and 38% (56 out of 148). The average proportion of geographic data was 86% and the median 90%. In terms of GIS-ready datasets, there were two small municipalities with four and ten datasets, none of which were GIS-ready. These were followed by four large cities with 7% to 13% of geographic data being GIS-ready. However, the majority of catalogs contained a significant proportion of GIS-ready data, with an average of 71% and a median of 89%. In fact, in 13 of the 46 municipalities, all geographic data (100%) were offered in GIS-ready format!

Maps, Platforms, and Approach

Another important observation made was the software platform used and the presence of maps in conjunction with the open data catalogs. The majority of platforms used were basic HTML pages supporting direct download. The remaining catalogs used the following platforms: Open Government Data Initiative (OGDI), CKAN, Socrata, and Esri’s ArcGIS Open Data.

The HTML-based direct download platform was not consistent in the use of maps and often directed users to separate mapping Web sites. Examples of this approach include Toronto, Ontario, and Victoria, British Columbia. There also were instances of maps in PDF files, the use of image formats, and links to Bing and Google maps.

The CKAN platform was consistently absent of any maps, as seen in Montreal, Quebec, and Surrey, British Columbia. OGDI was not consistent, for some catalogs like that of the city of Waterloo, Ontario, did possess maps, while Red Deer, Alberta, did not. OGDI was consistent in providing spreadsheets of datasets, but differed in terms of use of its graphical capabilities.

The Socrata platform had the most extensive GIS-like features of the five identified platforms. Edmonton, Alberta, and Langley, British Columbia, are examples of catalogs that use this platform. Users are capable of collecting datasets as a type of layering, opacity was an option, and the ability to add or change base maps was

Figure 1. Overview of municipal open data catalogs surveyed across Canada, including 23 catalogs surveyed in November of 2012 by Currie (2013), and nine (11) and 46 catalogs surveyed by the authors in December of 2013 and January of 2015
possible. The ability to create graphs in Socrata was fairly easy because of the many graph-making options for users.

The Halifax, Nova Scotia, open data catalog was the lone user of Esri's ArcGIS Open Data platform. This platform offers simple maps, access to a spreadsheet version of the data, as well as the possibility to create graphs.

The majority of catalogs have well-maintained glossaries. This is important as file formats and context are provided to assist the user, and often are accompanied by a question-and-answer section. If the availability of GIS-ready files is to increase, it will be beneficial to guarantee that these formats are well explained.

The presence of maps to visualize the user's relationship to the location of a dataset is essential for transparency and engagement. The display of a map immediately with the data creates a strong impression for the user to understand the context of the data. With a strong relationship established with the data, the excitement of engagement can turn into map making, the creation of apps, and research.

**DISCUSSION: MEDIATING OPEN GEOGRAPHIC DATA**

**Geographic Digital Literacy**

An important question for traditional open data is who is benefiting? For users to benefit from open data, they must have technical access, the ability to understand what they are looking at, as well as the ability to share what they find, to achieve government transparency (Jaeger and Bertot 2010). To truly benefit from open data, digital literacy, i.e., a basic understanding of information technology, is required, alongside the ability to interpret the data and metadata (Scassa and Singh 2015). Digital literacy also can enable open data users to create apps, reports, tools, and visualizations such as maps. In the case that governments aim to engage researchers, entrepreneurs, and citizens with open data technologies, an understanding of

<table>
<thead>
<tr>
<th>City or Regional Municipality</th>
<th>2012</th>
<th>2013</th>
<th>2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halifax, NS</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Fredericton, NB</td>
<td>x</td>
<td>x</td>
<td></td>
</tr>
<tr>
<td>Gatineau, QC</td>
<td>x</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Montreal, QC</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Quebec City, QC</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Sherbrooke, QC</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Burlington, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Guelph, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Hamilton, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Kitchener, ON</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>London, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Mississauga, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Niagara Falls, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Niagara Region, ON</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Ottawa, ON</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Region of Peel, ON</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Toronto, ON</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>City of Waterloo, ON</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Region of Waterloo, ON</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Welland, ON</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Windsor, ON</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>York Region, ON</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Brandon, MB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Winnipeg, MB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Regina, SK</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Saskatoon, SK</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Banff, AB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Calgary, AB</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Edmonton, AB</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Grande Prairie, AB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Medicine Hat, AB</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Red Deer, AB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Strathcona County, AB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>County of Grande Prairie, AB</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Chilliwack, BC</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Langley, BC</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Nanaimo, BC</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Regional District of Central Okanagan, BC</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Regional District of North Okanagan, BC</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Prince George, BC</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Saanich, BC</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Surrey, BC</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Vancouver, BC</td>
<td>x</td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>District of North Vancouver, BC</td>
<td>x</td>
<td></td>
<td>x</td>
</tr>
<tr>
<td>Victoria, BC</td>
<td></td>
<td>x</td>
<td>x</td>
</tr>
<tr>
<td>Whitehorse, YT</td>
<td></td>
<td></td>
<td>x</td>
</tr>
<tr>
<td></td>
<td>23</td>
<td>11</td>
<td>46</td>
</tr>
</tbody>
</table>
how these stakeholders would embrace the data and associated technologies needs to be developed (Zuiderwijk et al. 2015).

Arguably, GIS-ready datasets require professional expertise to be used effectively, thus catalogs including large proportions of GIS-ready datasets, up to 100% of all geospatial content, cater to this group of potential users. Conversely, some catalogs offer few GIS-ready data for download while they still include a large amount of geographic data. However, these are presented in map form rather than as machine-readable data. The city of Edmonton’s catalog with 442 geographic datasets out of a total of 521 open datasets is the prime example of this approach, as it provided only 32 of the 442 datasets (7%) in GIS-ready format.

The slower growth of GIS-ready data compared to all open data noted previously can be explained from both the supply and demand side. Being data experts, it is likely that staff in municipal GIS departments were among the first to be tasked with providing open data and found it to be most efficient to simply upload existing datasets, such as Shapefiles, “as is.” On the other hand, with the growth in popularity of using open data among non-GIS experts and the advent of easy-to-use online geocoding and mapping tools such as CartoDB and MapBox, there may have been an increasing demand for raw data tables with place names, postcodes, or x/y coordinate locations rather than GIS-ready data.

A promising approach to standardizing geospatial open data and promoting their reuse is demonstrated by the European Union with its 2003 Public Sector Information directive and 2007 Infrastructure for Spatial Information in the European Community (INSPIRE) directive, as discussed by van Loenen and Grothe (2014). Ideally, all geographic data should be provided simultaneously in raw form, in GIS-ready format, and visualized in predefined maps within the catalog. This would satisfy the desire of both those who are well educated in GIS as well as those technically savvy outside the GIS community to quickly engage with the data. Predefined maps would assist all potential users in comprehending the data. These maps should be placed within the same window of the catalog to avoid confusion and to quickly grasp the user’s interest.

**Geographic Open Data Infrastructure**

In Canada, the Honorable Tony Clement has equated open data to a natural resource, which should be profited on, making the argument that it could promote innovation, economic development, and entrepreneurial applications (Sieber and Johnson 2015). In 2010, individual departments of the federal government partnered with municipal, provincial, and territorial governments on their open data initiatives, providing mostly raw data with a geographic context (Davies and Lithwick 2010a). Furthermore, as of 2010, most provincial governments across Canada had made the move to establish interactive online portals, with the exception of Quebec and British Columbia (both of which now have open data portals). Municipal governments had introduced test versions of “open data” portals, where raw data were provided to the public in an accessible format with individual “terms of use” and typically in an A-Z directory listing of datasets (Davies and Lithwick 2010a).

As of 2016, there were 98 open data catalogs in Canada, spanning provincial, municipal, and local jurisdictions (McKinney 2016). The continued growth of the open data movement has formed a civic community, and this community should include those who manage the catalogs. Dialogue with, and between, catalog managers would promote uniformity of types of datasets published regardless of geospatial content. This would further enable research, and app and map development across the country. For example, an application that tracks wait times for public transit in Toronto could be more easily replicated for Montreal.

**CONCLUSIONS**

The municipal open data movement in Canada seems to have grown organically, resulting in a wide variety of approaches to mediating open geographic data. The delivery of geospatial contents within Canada’s municipal open data catalogs has...
increased significantly since the baseline survey by Currie (2013) in November of 2012. Our December 2013 and January 2015 scans establish a substantial record for the analysis of how the open data movement has developed and can serve as a basis for future comparisons. Based on the general prevalence of geographic references in data-centered decision making, we expect the proportion of geographic data to remain near the 80% level. The proportion that is offered in a GIS-ready format will probably be dictated by constraints on the supply side and by changes on the demand side, including the future development of GIS technology. If GIS becomes more ubiquitous through cloud-based, easy-to-use apps, it will be very important that geographers, planners, and environmental scientists continue to engage with the open data movement to demonstrate best-practice uses of open data.

This study merely provides a snapshot of open data in one country at one (or two) specific points in time. Some decisions regarding the categorization of geographic data and GIS-ready data formats may have been subjective and slightly influenced the results. For example, at the time of writing, the increased importance of the GeoJSON format and data access through APIs should be taken into account. The 80% proportion of geographic data among the surveyed open data cannot prove the prevalence of geographic data among all data, for open data may have other characteristics that influence this proportion, not the least being created for government and public administration purposes, which may require a geographic perspective that data-driven planning and decision making in the private sector may not entail. Nevertheless, we are confident that the order of magnitude of the reported results and the general trends over time can be reproduced and will continue for the foreseeable future.

**About the Authors**

**Edgar Baculi,** a geographer, GIS professional, and map enthusiast, is a consumer data administrator for Consumer Vision. After completing a BA in geographic analysis at Ryerson University in 2016, he worked as a geocoder and transit reviewer for RA Malatest & Associates. He conducted the December 2013 and January 2015 surveys of Canadian municipal open data for this article during his undergraduate studies. He will be returning to Ryerson University starting in the fall of 2017 to complete the Master of Spatial Analysis (MSA) degree.

**Dr. Victoria Fast** is an assistant professor in the Department of Geography at the University of Calgary. Specializing in urban GIS, her research profile spans participatory GIS, VGI, open data, and smart cities. She employs Web-based collaborative technologies—including open data—to start conversations, build capacity, and strengthen resilience between government and citizens.

**Dr. Claus Rinner** is a professor and chair of the Department of Geography and Environmental Studies at Ryerson University. His research expertise within geographic information science covers geovisualization, participatory GIS, and spatial decision support. More recently, he became interested in open geospatial data, as exemplified in this article, and in 3-D–printed geographies.

**Corresponding Address:**

Ryerson University  
Department of Geography and Environmental Studies  
350 Victoria Street  
Toronto ON M5B 2K3 Canada  
crinner@ryerson.ca
References


Currie, L. J. 2013. The role of Canadian municipal open data initiatives: A multicity evaluation. Queen's University.

Davies, A., and D. Lithwick, 2010a. Background paper government 2.0 and access to information: 1. recent developments in proactive disclosure and open data in Canada. Publication No. 2010-14-E.


INTRODUCTION

Opening up civic data has become part of the modern infrastructure of municipal governments. Opening data has several goals: To enhance transparency and accountability, increase government efficiency and service delivery, and promote economic development and business intelligence (Sieber and Johnson 2015). Most open data provisions, however, consist of little more than “throwing data over the wall” (ibid). This may ensure open data that can be “freely used, modified, and shared by anyone for any purpose” (Open Knowledge International 2012) begins to meet its goals, but can fail to describe all the actors involved in open data adoption, sustainability, and value. Regardless, government remains confident in the potential of open data apps to improve public services (Bates, 2014).

Concepts such as crowdsourcing (Brabham 2009), citizen science (Haklay 2013), and Volunteered Geographic Information (VGI) (Goodchild 2007) offer new sources of data for governments and introduce opportunities for citizen-side collaboration. Indeed, governments look to citizen contributions of data as a form of public engagement and a substitute for declining government resources.

Regardless of the source (government or citizen), civic data will require infomediaries. This has become the accepted norm in public use of open data, by which “entrepreneurial actors will create ‘apps’ that make data accessible to citizens” (Davies, Perini, and Alonso 2013, 14), which also can extend to government uses of data.

The term infomediary originally was defined by Hagel and Rayport (1997) as “custodians, agents, and brokers of customer information, marketing it to businesses on consumers’ behalf.” Infomediaries (sometimes referred to simply as intermediaries) now are generally understood as actors that help manage data and information between a data source and user (Janssen and Zuiderwijk 2014). They partake in converting data to information, such as creating databases that track activities of local politicians to increase citizen engagement with politics, or promoting data interoperability by combining local-level data as a form of community building (Worthy 2015). We use an Actor-Network Theory (ANT) definition of the actor, which is anything that “acts or to which activity is granted by others” (Latour 1996, 373). This allows us to use the “actor” to describe any human or non-human entity in the open data app ecosystem that exerts control over data. We define infomediaries as actors that transform or control data, but exist outside of government control. Not all actors are infomediaries; decision makers in government who exert their own control over data are not infomediaries.

Traditionally, government has outsourced much of its information technology (IT), some of which is developed by traditional sources such as private-sector consultants and, more recently, created by emergent sources such as participants in hackathons (Chen and Gant 2001, Johnson and Robinson 2014). Infomediaries market their services with easy-to-use public interfaces and back-end analytics. The current Web 2.0 paradigm, defined by O’Reilly (2007) as a shift towards dynamic and user-generated web content, treats data and software as a dynamic service, making outsourcing of analytics and hosting to infomediaries more attractive. Government must trust that these infomediary services remain operational so that data continues to flow. Infomediaries may exert control in the form of data (re)formatting, aggregation, interface design, analytics, and mapping. Control includes access to data, as access can allow or limit the potential to manipulate data. Infomediaries may be human (e.g., software developer) or
non-human, (e.g., data server). We are interested in identifying the significance of all actors because infomediaries presumably augment raw data with operational functionality and value.

In this article, we describe infomediaries found in five municipal applications (apps) that are built around civic data, using ANT to frame them within the app ecosystem. An ecosystem view is important as it does not restrict us to just individuals as a unit of analysis, but allows us to comment on the broader sociotechnical environment they inhabit. We focus on apps because civic data is likely to be offered in a mediated form, including crowdsourcing sites and citizen dashboards (Sandoval-Almazan, Gil-Garcia, Luna-Reyes, Luna, and Rojas-Romeró 2012). First, a review of government models of service provision afforded by Web 2.0 allows us to define the role of the infomediary. Then, we cover our methods of app selection and interviews with respondents with direct experience in appifying the raw civic data. We conclude with the increased significance of the infomediary as part of the government public service delivery process.

THE ROLE OF INFOMEDIARIES IN THE CIVIC DATA APP ECOSYSTEM

The rise of infomediaries is driven by two congruent forces, one from government and one from the information technology industry. For government, it is a matter of practicalities and paradigms. Governments engage external parties for numerous reasons, among them to draw upon global talent, leverage economies of scale, free up resources, enhance flexibility and responsiveness to new conditions, limit liability, and compensate for resource constraints (Grimshaw, Vincent, and Willmott 2002; Stroh and Treehuboff 2003). Governance paradigms such as New Public Management (NPM) suggest a structural shift in the relationship between government and other sectors and civil society. Here, government becomes a manager of public service providers, where its functions and responsibilities are preferably delegated to external parties (Denhardt and Denhardt 2000, Hood 1995).

Historically, this outsourcing has predominantly been the domain of private-sector interests. Outsourcing has broadened to include citizen-sourcing (Nam 2012), where citizens-at-large actively contribute to coproduce public goods and services (Ostrom 1996). In the Government 2.0 paradigm, envisioned by O’Reilly (2011), the sector of the service provider does not matter so long as government operates as a platform upon which others can innovate. Cities, for example, should function solely to provide the resources, rules, and regulations to govern public services. External parties will do the rest.

O’Reilly (ibid.) views government data as the newest infrastructure to be outsourced. He and others (e.g., Klischewski 2010) argue that if cities simply open data then outside brokers will emerge to deliver services, whether filling potholes or constructing apps. This potential is enabled by Web 2.0 technologies, which serve the long tail of users, provide a rich user experience, and enable software as a service (SaaS) of multiple interoperable components, and where citizen-generated information is as important or more important than official data (O’Reilly 2007).

Web 2.0 aids and complicates government functioning. Adaption to new innovations can be hindered by rules and regulations governing bureaucracy. Adoption of Web 2.0 tools, like harvesting of social media, results in higher data volumes, new skills and expertise, and higher labour and IT maintenance costs. Government employees may find themselves subject to information overload (Sivarajah, Irani, and Weerakkody 2015) simultaneously as new tools potentially disrupt organizational hierarchies (Bond 2015, Klischewski 2010). Relying on such Web 2.0 tools may prevent government from carrying out some of its basic required functions. For example, archiving communications is almost impossible when they sit on social media platforms outside of government control, designed to be ephemeral and subject to business models that may terminate a service at will (Bertot, Jaeger, and Hansen 2012). Government could see the potential “loss of ownership control and authenticity of the final products” of communications or data (Freeman and Loo 2009, 77) as they are unable to control outcomes.

With Web 2.0, government becomes dependent on a multiplicity of third parties, technologies, and standards to deliver services from data collection to data distribution. Web 2.0 shifts app development from large, customized, and single-sourced turnkey projects to modular platforms supported by many developers. Government dependence on proprietary standards and platforms has been a vulnerability (Evans and Reddy 2002); the shift to Web 2.0 and modular software promised freedom from singular platforms (Fishenden and Thompson 2013). However, a platform with many complex and interdependent software components, accompanied by high levels of expert labour, can expand reliance on external parties (Schneider and Sunyaev 2016). Yu (2014) describes this scenario as mutual dependence. The fewer the alternatives available to government, the higher its dependence on an infomediary. He notes that high mutual dependence can result in inflexibility and loss of control over data. Given sufficient trust, however, mutual dependence fosters mutual investment and can strengthen long-term public-private partnerships. The complications introduced by outsourcing suggests the notion of the “government as a platform” may itself be counterproductive to the aims of government.

METHODS

We tracked the flow of civic data (open data and citizen-sourced data) to identify the actors involved in the use of civic open data. Rather than choosing specific datasets or cities as units of analysis and examining the infomediaries of a government’s open data portal, we chose to explore infomediaries by which open data is expressed to the public via an application. We selected apps that represented the best examples of app development at the municipal level in Canada. To identify the actors, some existing in government and some existing outside (i.e., infomediaries), we conducted qualitative interviews of app developers and government officials. Identification of apps and actors was assisted by examination of public information from government websites. Based on qualitative interview data and information from their...
respective government websites, we traced the data from source to user, and revealed actors along the way.

App Selection

Sampling of apps was based on a typology of citizen and market-oriented apps (Sangiambut and Sieber 2016). This typology helped us choose apps based on directionality of data flow. First, a scan of existing open data apps was performed in Canada, based on official app listings from Canada’s major open data cities’ websites: Edmonton, Montreal, Ottawa, Toronto, and Vancouver. The cities of Edmonton, Ottawa, Toronto, and Vancouver comprise the initial network of open data cities in Canada, known as the G4 (Carl 2012). As such, these cities were some of the leaders in Canadian open data production, had the longest sustained production of open data, and likely the best examples of local app development. Best cases were selected based on several criteria: Whether open data flowed to or from government (apps from both directions of data flow were chosen), popularity (number of users), and currency (how well the app was maintained). Apps that appeared to be “dead” (no patches in over six months) were not considered. Sampling was restricted to the local Canadian municipal level to remove the mixing of levels of government because municipal, regional, and federal governments can have differing operations and compositions as well as motivations for releasing datasets. Data flowing to government had to be used by government, otherwise there was no mediation. Apps outsourced by government were the best guarantee that government was involved in data production or collection. Ultimately, we chose five Canadian municipal-level open data apps: Citizen Dashboard, Ottawa Transit, Citizen Budget, Toronto Cycling App, and VanConnect.

Citizen Dashboard was chosen as it won a 2014 Public Sector Leadership Award for the (at the time) progressive nature of this app (Quigg 2014), and because it was a directly outsourced government app. Ottawa Transit was selected because it won an award at the Apps4Ottawa open data hackathon, had active development (bimonthly updates), and a download base of more than 20,000. Citizen Budget was chosen because it was the only public consultation app currently in use. Its developer, Open North, has demonstrated its use with clients in multiple cities across Canada and in the United States of America. Toronto Cycling App was chosen because it is an app outsourced to the private sector, its relatively high user base of more than 5,000, and its direct usage in city infrastructure planning. In particular, it was advertised as a tool being used to inform planners in their next cycling infrastructure plan, presenting a direct link to government. Finally, while VanConnect had strong developer support and a large install base of over 10,000 users, it also was included because it had both citizen reporting and push notification functionality, which exemplified hybrid or bidirectional data flow. This suggested the possibility of additional dependency relationships between government and developer to be observed.

Semi-structured Interviews

To ascertain the actors in the app ecosystem, semi-structured interviews were conducted with app developers and government officials in managerial or leadership positions. Semi-structured interviews could generate more nuanced responses on the dependencies between government and outsourced developers and any infomediaries they use (such as their own software platforms or third-party Application Programming Interfaces [APIs]). Semi-structured interviews also allowed us to ask probing questions on the reasoning behind decisions and their origins, which increased the likelihood of identifying additional actors that were unknown to the respondent. By selecting those who were in leadership positions related to the app, we were able to ensure respondents had a high level of historical and working knowledge on the app and data. At least two—one individual from each app’s respective government and developer organizations—respondents were interviewed. If a respondent could not answer a question, we asked him or her to refer to a colleague in his or her organization. A total of 13 individuals were interviewed. A developer (PublicStuff) provided two respondents in the same interview, and the City of Toronto and City of Vancouver each provided two respondents in two separate interviews. Respondents were asked to describe the app and data, its development, and their interactions with their government or developer counterparts. They also were asked their views on outsourcing and public services.

Mapping of the Actors

To map the app ecosystem and find its obvious and not-so-obvious actors, we used two methods. First, we followed the data to determine the human actors. This was a snowball sampling method to identify interview respondents connected to the app and data flow, based on a similar implementation by Davies and Frank (2013) and Sands et al. (2012). We adapted this by selecting apps as a starting point, then looking forwards and backwards to identify who to interview at the source and the destination of the data flow. Interviewees were asked to identify the next person down the data flow to interview.

Following the path through which data travels, from source to destination, captured the technical environment of the app. However, it did not capture the influence of non-humans or influences from entities outside of the data flow ecosystem. For instance, certain non-human actors may exert influence over the flow without the data physically passing through them, such as open data licences. To address this, we utilize Actor-Network Theory (ANT) to frame the interview data within an app ecosystem, with a view that an actor is any entity that “acts or to which activity is granted by others” (Latour 1996, 373), whether human or non-human. Callon states that examining power relationships entails “describing the way in which actors are defined, associated and simultaneously obliged to remain faithful to their alliances” (Callon 1986, 19). ANT therefore allows one to examine power relations and the conditions for power flows. Actors such as software platforms and APIs, and actors seemingly disassociated
with data flow such as government legislation or directives, thus could be important in shaping the relationship between government and infomediaries.

The use of ANT in government IT projects is not new. Stanforth (2003) has examined actors’ power over the development of public expenditure information systems in Sri Lanka, while Walsham and Sanjay (1999) examined GIS technology’s embodiment of Western values in India. Heeks and Stanforth (2007) state it is common to consider ANT as a broad approach or perspective. They note that researchers have chosen specific aspects of ANT, such as the four moments of translation (Callon 1986), as a lens rather than a comprehensive analytical method. Carroll, Richardson, and Whelan (2012) documented some basic steps of adopting an ANT approach to include: Identifying actors, their relationships, tracing of actions (what activities led to these relationships), and which actors enable or inhibit certain actions within the network. Our semi-structured interviews were geared towards answering such questions.

Respondents problematized the ecosystem and identified actors by describing the origins of the app, key decision makers behind its origins, and current interactions between app developer and government officials. Their descriptions (e.g., who made decisions, what technologies they relied on) allowed us to see which actors had power over data and over other actors (via data). When individuals (such as corporate-level city officials) had enough power over data to differentiate their actions from that of their organizations (i.e., they expressed their own agency), they were considered unique actors within the ecosystem and became fixed. However, if an individual was beholden to their organization’s goals, he or she was subsumed into a larger actor.

**CASE OVERVIEW AND DATA FLOW**

In the following sections, we build up the app ecosystem. First, we give an overview of each app ecosystem’s actors. Then we present results describing the relationship between infomediaries and government, and the types of control over data they wield. We concentrate on the infomediaries in the ecosystem, but chose not to ignore government actors because we observed numerous possibilities for infomediaries to enter at different stages of data flow before data exits government control.

**Citizen Dashboard**

Citizen Dashboard is a browser-based interactive app from the City of Edmonton (see Figure 1). The app displays 64 (as of June 2016) performance measures that inform residents on progress towards city goals. Measures include event-specific measures such as snowploughing, monthly measures such as Disabled Adult Transit Service (DATS) on-time performance, and annual measures such as city operations greenhouse gas–emissions targets. Behind every performance measure is a continually updated dataset located in the City’s open data catalogue. Citizen Dashboard’s development, maintenance, and hosting were outsourced to a U.S.–based private-sector developer—Socrata. Socrata sells this dashboard to other customers under the name Open Performance.

Data flows from government to citizen primarily through Socrata. City officials in all branches of the administration are responsible for updating the data behind each performance measure. An employee in the Corporate Strategic Services department oversees all updates to Citizen Dashboard’s performance measures to ensure they are conducted in a timely manner. Only employees in Corporate Strategic Services have edit permissions for the performance measures’ descriptive texts and visualizations,
which places them in a position of power. Performance measure datasets are manually updated and stored in the city’s open data catalogue; all hosted by Socrata. The data visualization in Citizen Dashboard then is automatically updated by querying the open data catalogue via the Socrata Open Data Application Programming Interface (SODA).

In these cases, the data is not necessarily open throughout the ecosystem. Only when they are uploaded to Edmonton’s open data catalogue does the data become open.

Ottawa Transit

Ottawa Transit allows users to plan trips across Ottawa and provides real-time notifications on bus arrivals (see Figure 2). Ottawa’s transit agency, OC Transpo, uses a Computer Aided Dispatch and Automated Vehicle Location (CAD/AVL) software platform, called Hastus (developed by Giro), to create bus schedules. Hastus automates much of the data creation (employees select input variables but the software performs calculations) and transforms bus schedules into the General Transit Feed Specification (GTFS). In addition, transit schedule planning is dynamic, responding weekly or daily to changes in demand. Hastus also has an AVL component that can receive GPS sensor data and track bus locations in real time. This data is delivered to the public via an inhouse API called Live Next Bus, which is entirely automated.

Ottawa Transit was developed by 3lywa Solutions, a one-person firm based in Ottawa. App development was instigated by the developer’s participation in a city hackathon. The app uses both the static bus schedule and the streaming bus locations to aid users in navigation. The developer downloads the entire bus schedule every season and transforms the GTFS-formatted data into a single relational database. Bus location data is accessed directly by the app in real time. OC Transpo provides guidance on the technical requirements for the data, such as how to access the API, but imposes no creative limits on developers.

A third-party API token management service (3Scale) manages API calls for Live Next Bus by sitting in between app developers and OC Transpo servers. It serves as a potential block for data as it forwards only authorized calls and establishes quotas on queries. The terms of use for both the GTFS and the Live Next Bus data are covered by an open data licence. Authentication was instituted because OC Transpo, prior to this app, had denied access to another infomediary that had been scraping schedule data from its website.

Citizen Budget

Citizen Budget is a browser-based app that collects citizen feedback on the Borough of Plateau-Mont-Royal annual budget (see Figure 3). The Borough of Plateau-Mont-Royal is part of the City of Montreal but possesses separate jurisdictional and data-collection authority. The app provides an interactive form within which citizens can act as finance officers and simulate budget allocation. Residents are asked whether they would like to increase or decrease spending and taxes on municipal operations, such as snowploughing, and whether they would like to initiate capital projects such as a new library. The app reflects the real funds available to the Borough and can constrain users to create a balanced budget. Budget data is available as a CSV file or in an automated PDF report with data charts.

Data in the Citizen Budget ecosystem flows through few actors. To create the questions in the app that guide users through the budget, political aides collect initial data from the City of Montreal’s various departments and relays this to Open North via an electronic form. Open North hosts the online dashboard and the citizen-sourced data on its own servers. The data then
passes to the Plateau-Mont-Royal administration. Political aides within the administration interpret both the automated report and the raw tabulated data and report their findings to the Borough Mayor and councillors for their budget planning. Aggregated data is revealed at the annual budget consultation to support budget decisions. Data also is only open once it reaches the public budget consultation, and is publicly available only in aggregated form.

The app is a product offered by Open North, a Montreal-based non-profit organization that promotes democratic participation through online tools. At the time of surveying (2015), the app was not in use because of a restructuring of the city’s budget allocation.

**Toronto Cycling App**

Toronto Cycling App collects cyclists’ travel routes to assist the City of Toronto’s Cycling Infrastructure and Programs division plan Toronto’s cycling network infrastructure (see Figure 4). App users contribute their cycling trips (GPS coordinates) and can complete an optional survey of demographic information and cycling habits. The app can display cycling-related information from Toronto’s open data catalogue on a map.

The app was outsourced under a Request for Proposal bidding process to a private sector Ontario-based developer called Brisk Synergies. After individuals contribute information, the app sends the data to Brisk Synergies’ data server. Here it is error-corrected, converted into line segments, and attached to the road network. Brisk Synergies hosts this data and provides an online portal for
city officials to visualize the data and download for further analysis in their own systems.

Toronto planners input the data into their own geographic information system (GIS) (Esri’s ArcGIS software) for spatial analysis. The output then is used in cycling infrastructure planning to map demand for routes. It was used to inform the 2015 Cycling Network Plan, an extensive five-year cycling infrastructure plan proposed to the council.

Planners in Cycling and Infrastructure Programs are the primary users of the data, although other employees from the city’s Transportation Services division also can access the data. The City of Toronto respondent also mentioned a potential conversion of aggregated data into open data for Toronto's open data catalogue. When this is enacted, it will offer another point of data transmission.

VanConnect
VanConnect is an app used to file service requests akin to calling a 311 telephone hotline for non-emergencies (see Figure 5). Using either a mobile app or browser interface, users fill in a form to categorize the issue (e.g., pothole, broken streetlight, graffiti), map the location, and add any other information required. The objective of the app is to provide an alternate channel of communication for service requests and facilitate more automation of the process. VanConnect is developed by PublicStuff, a U.S.-based company and subsidiary of Accela.

The app sends the collected service requests to PublicStuff servers, which then forwards them to the City of Vancouver 311 Contact Centre's Customer Relationship Management (CRM) tool called LAGAN. The CRM automatically forwards the service request to a City of Vancouver department or an external organization such as Canada Post if the request lies outside the City of Vancouver’s jurisdiction. The Director of Digital and Contact Services is a key decision maker who manages 311 Contact Centre operations; whereas the City’s Digital Strategy, a development framework, dictates the Centre’s motivation for automation and digitization.

PublicStuff’s online dashboard is accessible by 311 Contact Centre employees and allows government to modify the app itself, such as adding “widgets” or modules that display information and maps and send push notifications. This means that data flows both to and from government. If users register their contact details, the city can send automated responses (via PublicStuff), notifying users of the status of their service requests. Finally, all 311 service requests (including telephone and e-mail) are aggregated and automatically published on Vancouver’s open data catalogue.

PERCEPTIONS OF RELATIONSHIPS OF INFOMEDIARIES IN AN OPEN DATA ECOSYSTEM

This section covers responses to questions about the relationships between human infomediaries and government. Government respondents for all apps were unanimous in their view that external actors were essential to fill resource gaps in municipalities. The City of Toronto’s Cycling and Infrastructure Programs is an analytical and planning unit, but reportedly lacked the resources to develop its own software. The City of Vancouver respondent stated that developing VanConnect “was not even on our radar to do it inhouse.” Regarding their prior internal projects, “version control is very poor, it doesn’t seem to grow very quickly, and usually the talent that built it, when they move on, we lose a whole bunch of knowledge.” A third-party developer, therefore, is necessary to ensure a project will endure.

The Manager of Transit IT at OC Transpo’s message was
it is infeasible and irresponsible for government to maintain the application because government cannot and should not keep pace with technological innovations. “The list is endless in terms of what [software] you have to support. The cost of that cannot be borne by the taxpayer; unfortunately, that’s the reality.” For this reason, OC Transpo only developed an app for the iOS mobile platform and leaves it to third parties to develop apps for other platforms.

Socrata’s responses were contextualized in the ideology of open government. Socrata argued that “access to information is/will change the way governments operate, both internally and externally.” Socrata views its infomediary role as promoter of a larger open government movement in which government data needs to be freed for repurposing. It is business model reinforced by ideology:
no it [data content] doesn't really matter [to Socrata]. But because we are also concerned with consumption . . . we try to encourage the data to be more rich . . . The more high value datasets individual customers publish, the more useful it is both to the citizens and to the developer community.

Socrata’s survival depends on governments publishing quality open data, which suggests a symbiotic relationship with government. Other developers expressed much less codependency with government. PublicStuff’s respondent expressed no opinion on data once it entered the government ecosystem. When asked how they hoped VanConnect’s service request data would be used by the city, they separated their role from government, where “our hope is that the city will use the data to respond to citizens requests for citizens, but really that’s not for us. We provide a service for clients, they can do whatever they want with the data.”

Infomediary developers were driven by varied motivations. Open North is Canada’s largest open data non-profit. Its mission is to promote democratic participation in more transparent and accountable government. Open North’s respondent saw its role as guide or mentor to government in the implementation of Citizen Budget. An advisory role was necessary as it was the Borough’s first attempt at an online tool. The respondent noted that, “usually they [Plateau-Mont-Royal] don’t feel comfortable enough, so they always want for someone to say ‘yay that’s good’” and that the Borough did not always know what it wanted from Citizen Budget. Developers, such as Open North and Socrata, may view their roles in the ecosystem as greater than a single government contract. This differs from traditional project outsourcing. Infomediaries may be distinguished from traditional outsourcers in part because they integrate into outsourcing an intention to ensure civic data meets its objectives of citizen engagement and accountability.

Finally, the nature of civic data may reveal personal motivations driving app development. For the developer of Ottawa Transit (Mlywa Solutions), “This whole application came from a personal struggle with busing, [which] started because of me as a user. I wanted to get data.” The developer saw a gap in quality transit apps, particularly on the Android platform, and decided to create one for the developer and the community. The developer therefore positioned him or herself as both private-sector infomediary and representative of civil society. Our research finds the sectors of the developers start to blur, as do the boundaries between government and developer.

Table 2. Definitions of types of control

<table>
<thead>
<tr>
<th>Types of Control</th>
<th>Definition</th>
<th>Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data collection</td>
<td>Data creation or acquisition</td>
<td>City departments, stand-alone software</td>
</tr>
<tr>
<td>Data access</td>
<td>Authority to download, stream, or view data and at what volume/rate.</td>
<td>Data server, tokens and authentication, online data dashboard, open data catalogue</td>
</tr>
<tr>
<td>Data storage</td>
<td>Container or custodian for data</td>
<td>Data server</td>
</tr>
<tr>
<td>Data structuration</td>
<td>Methods of categorization and classification</td>
<td>Developer, APIs, and data schema</td>
</tr>
<tr>
<td>Data modification</td>
<td>Changes (e.g., arithmetic operations, geometric transformations, conversions, conversions) of data or data fields</td>
<td>City departments</td>
</tr>
<tr>
<td>Representational transformation</td>
<td>Changes in how the data is represented or interacted with.</td>
<td>Online data dashboard, APIs, and data schema</td>
</tr>
<tr>
<td>Data mandate</td>
<td>Authority, ordinances to create, use data. Includes licences, legislation, mission statements</td>
<td>Data directives</td>
</tr>
<tr>
<td>Data documentation</td>
<td>Data about data, which explains, e.g., provenance, purpose, time, location, and other descriptions. Includes metadata, technical specifications</td>
<td>APIs and data schema, online data dashboards</td>
</tr>
<tr>
<td>Data augmentation</td>
<td>Extensions of data. Includes alerts, comments about the data</td>
<td>Online data dashboard</td>
</tr>
<tr>
<td>Data literacy</td>
<td>Material (physical, virtual) to increase written comprehension, numeracy, and data utility</td>
<td>Developer</td>
</tr>
<tr>
<td>Data analysis</td>
<td>Numerical, statistical, or graphical treatment of the data</td>
<td>Developer, stand-alone software</td>
</tr>
<tr>
<td>Data advocate</td>
<td>Promotion of data sharing, open government, and open data</td>
<td>City departments, developer</td>
</tr>
</tbody>
</table>
ROLES OF GOVERNMENT AND INFOMEDIARIES IN THE OPEN DATA ECOSYSTEM

In this section, we categorize the actors in each app ecosystem. Table 1 categorizes actors according to their roles, their attributes, and their types of control over the data or ecosystem. This allows us to separate the non-human and human infomediaries that influence data flow. Categorizing infomediaries allows us to explore the variations in actor roles. Table 2 defines the forms of control that the actors can exert. As seen in Table 1, all but one kind of actor exerts multiple forms of control over data, and government does not wield all the forms of control over data found in the app ecosystem. Instead, infomediaries express overlapping or completely different forms of control over data, particularly if they create their own technologies (such as APIs). Breaking down the forms of control over data exerted by actors (particularly infomediaries) allows us to map the varying interdependent relationships between government and infomediary.

City Departments
City departments act as dominant control points in the ecosystem. They exhibit control over data access, collection, modification, and decision making. While not infomediaries themselves, they hold major influence in the interdependent relations throughout an ecosystem.

The City of Edmonton’s Corporate Strategic Services initiated Citizen Budget in partnership with Socrata and maintains administrative permissions for the app. All branches of administration are involved in data production but only one specific branch has oversight and the major interactions with the developer. Any changes, for example to data visualization or descriptive text, must obtain approval from Corporate Strategic Services. For Ottawa Transit, OC Transpo’s Manager of Transit IT (who also was an interview respondent) is responsible for more than information management and software procurement. The Manager of Transit IT took a hands-on role in app ecosystem by educating (building data literacy) and regulating (setting guidelines for) developers. The manager hosted developer hackathons to explain the data schema and how it could be used and explained API call limits. In this way, the respondent demonstrated the ability to act both as a representative of his or her organization and with his or her own agency.

Developer
Outsourced developers are the traditional infomediaries represented in the literature. We found these actors adopting important roles of data structuration, representational transformations, data analysis, data literacy, and data advocacy. In many cities, Socrata has been given control of the entirety of the open data portal, from publishing (access) to analytics (value). Brisk Synergies’ involvement in the collection and conversion of point data gives it control over the structure of data that is presented. Brisk Synergies also exerts control over the presentation and visualization of the route data via its online data dashboard. Skills such as processing point data, attaching line segments to the road network, and visualizing this data in an online interface were reported as outside the capacity of the City of Toronto’s cycling planners.

Developers also may take on important roles beyond simple outsourcing contracts. Open North acted in an advisory capacity to government in teaching borough employees and aids the potential of online consultation tools. In this way, the non-profit acted to improve data literacy and also advocate for open data and crowdsourcing.

Data Directives
City reports, terms of service, and data laws can impact the ecosystem. Open data licences, such as the City of Ottawa’s Open Data Licence, stipulate that the data may be freely used and modified, provided the source is acknowledged. For Citizen Dashboard, the City of Edmonton respondent cited The Way Ahead, a strategic plan that sets out development goals for the city for the year 2040, as an instigator of performance measures for the public and mandate for increased transparency. The City of Vancouver’s Digital Strategy mandates the automation and digitization of public services. Legal jurisdiction over data is also a factor in the ecosystem. While Socrata does not claim ownership over the open data it hosts, its data servers are physically located in the U.S.A., which subjects City of Edmonton open data to foreign laws such as the U.S. Patriot Act. This concern has been raised for cloud-computing services in the past (Zhou, Zhang, Xie, Qian, and Zhou 2010) and represents significant obstacles to the engagement of U.S. infomediaries by Canadian cities, which operate under stricter privacy laws than does the U.S. (Scassa, 2013).

Open data Catalogue
Open data catalogues are the primary access portals by which governments control access to open data and its representation. Socrata’s open data platform provides the user interface to the data, and displays the data on a map. The catalogue becomes the gateway for data downloads and the data representation (e.g. web mapping). When a data catalogue becomes the host for data displayed in an app, such as Citizen Dashboard, dependency on this actor increases. Citizen Dashboard’s entire functionality relies on the City’s Socrata-powered open data catalogue being functional.

Control over representation can be exerted through control over the type of data or data formats that the catalogue allows to be stored. Storing point data in a CSV file can reduce the potential for complex analyses compared to a Shapefile. Typical data portals display a list of datasets for the user to download wholesale. However, Socrata’s open data catalogue does not store the dataset in discrete files. Instead, data is stored in a database, free of file associations. When a user downloads a specific dataset, it triggers an API call to the catalogue that queries the database and returns the specified file for download. This allows Socrata-powered data
catalogues to export to a wide number of file formats, and allows users to make data queries straight into datasets without having to download them. This implementation does not mean that all data is available in all formats. For example, Edmonton’s Tree Species dataset is displayed on a map in the open data catalogue, but is only downloadable in tabulated (CSV) and markup language (JSON, XML, RDF, RSS) formats, with no true spatial data format as an export option even though the dataset itself contains coordinate data. This limits the ability of users to perform complex analysis or mashups with other spatial datasets. Limiting the available formats also limits potential reuse to only those with the skills with that particular type of data.

**Online Dashboard**

Online dashboards exhibit control over representational transformation, data access, documentation, and augmentation. Socrata’s platform allows the City of EDMONTON to upload and modify data, and modify the content and visualization of performance measures. In cases such as Toronto Cycling App and Citizen Budget, an online dashboard is the only means for government to access the data; whereas others, such as Socrata, provide alternate access points such as an API and stand-alone software. PublicStuff’s online dashboard also is used to add widgets to the VanConnect app, which can display additional information or notifications to users.

**Stand-alone Software**

Some government organizations relied on large software packages to perform tasks. These software packages exert control over data access, collection, modification, analysis, and decision making.

Several examples suggest the impact of software on the ecosystems of open data. The real-time bus-locations feed from OC Transpo’s Live Next Bus API is completely automated and thus entirely reliant on the HASTUS system. The City of Vancouver’s 311 Call Centre’s LAGAN CRM also automates the rerouting of VanConnect service requests to service providers to negate the need for human mediation. This CRM is the central actor into which all 311 service requests flow. Because this software is responsible for categorizing requests, it dictates how app users structure their service requests. Stand-alone software, such as ArcGIS software used by the City of Toronto, are also entry points for the influence of their owners (in this case, ESRI) into the ecosystem.

**Data Servers**

Data servers act as important control points over data access and storage. Data for Citizen Dashboard and for the City of Edmonton’s open data catalogue are remotely hosted by Socrata. The city’s open data catalogue is stored abroad in the U.S., and thus is subject to U.S. data surveillance laws. The Citizen Dashboard’s web pages also are served by Socrata’s data centre, which extends the actor’s power over a city’s transparency policy. Interestingly, the city official at the City of Edmonton stated an incorrect belief that the City hosted its own open data catalogue.

**APIS and Data Schema**

APIS and schema control the structure, access, documentation, and representational transformation of data. Socrata’s Open Data API provides an important example. The API mediates data by shaping the categories of data that are received by and sent to the Citizen Dashboard. The Socrata developer stated that the platform emphasised quantitative analytics. Qualitative data, particularly text, represents a secondary, and more difficult to aggregate and visualize, priority. This has resulted in certain visualization issues for the City of Edmonton, where qualitative metrics or event-based performance measures are difficult to depict. Even its credit rating measure, which is represented alphabetically (i.e., A++, A+, A, B, C, D) is insufficiently numerical and thus must be readapted for Socrata. The city also differs from Socrata in the temporal frequency of data updates. Edmonton produces performance measures called Lot Grading Inspections, which are conducted on an annual basis but cannot be displayed as a monthly trend as visualized by Socrata. Differences are not inherently negative but demonstrate the adjustments necessary for new IT and the control induced by any adaptation.

A subtler control point is GTFS, which is used in Ottawa Transit. GTFS remains the de facto standard for displaying transit data online. While it is an open specification where changes are agreed upon by the community, GTFS still is overseen by Google. Like the Google Maps API (used in Ottawa Transit, Toronto Cycling App, and VanConnect), GTFS has become an industry standard that serves as an entry point for Google into the app ecosystem and illustrates that technologies are not entirely under the control and ownership of infomediary developers.

**Authentication**

The Ottawa Transit case revealed the use of a token management service, 3Scale, as a control point over data access. By limiting access to data to those with authorized tokens, the token management platform allows OC Transpo to enforce quotas on API calls. OC Transpo has the ability to block an app should it contravene the terms of service and the open data licence. Control over data access is also a control over the potential for data transformation. For example, quotas on API calls can introduce issues of temporal accuracy and uncertainty in streaming data visualizations.

**DISCUSSION**

Infomediary interventions operate throughout the civic data production process. Interventions vary in size and extent, ranging from providing a schema to managing the entire data infrastructure. As expected, interview results confirmed support for outsourcing, and exploration of actors revealed the types of control that are being outsourced. Infomediary influence extended far into government as interview results confirm that sectoral boundaries are increasingly porous.
Infomediaries assumed roles traditionally held by government IT departments, a transferral welcomed by government. OC Transpo, for example, argued that government should not bear the responsibility of creating and maintaining apps, particularly due to the development costs that would be a needless burden on taxpayers. This extended to open data portals. Socrata assumed control of the entirety of Edmonton’s open data portal. It should be noted that when we began a larger research project in 2011, Canadian cities including Edmonton expressed antipathy towards corporate management of open data portals. Cities, such as Toronto and Montreal (Plateau-Mont-Royal), now frankly admit their need to outsource, primarily due to their lack of app development capacity. Brisk Synergies and Open North therefore assumed government roles of data preprocessing (e.g., data aggregation and accuracy checking) and visualization. Potential complications, such as lack of control over the archiving or storage of data, were dismissed as government respondents were confident in the strength of contractual agreements.

Governments may delegate to infomediaries functions such as data publishing and distribution. The concern is that this will further a neoliberal delegation of power and responsibility (Leszczynski 2012). Coupled with cloud computing such as SaaS and storage services, government may be reduced or disassembled as it relinquishes control over more functions, a movement towards Government 2.0. Ceding control of functions, for example through the structuration of qualitative data, also suggests cities yield to a business model that effectively homogenizes cities. Contrary to the nuanced empowering potential of Web 2.0 (e.g., an enabler of multimedia narratives from citizens, ibid.), interdependency in the ecosystem can be quite limiting.

The level of dependence on infomediaries, particularly those embedded in the data production process, indicates a lack of substitutability for those services. If an outsourced service disappears, then government will need to adjust procedures that already were adapted to the prior service; services that may have been reified through policies and rules. Even if they chose a different vendor, government may require nontrivial reinvestment in hardware, software, and skills retraining. Efficiencies may be gained in shifts to new developers; however, this may be an irreversible process if government fails to retain the skilled labour and expertise required to rebuild data infrastructure. In such a case, government may lose the ability to evaluate outsourcing projects (Grimshaw et al. 2002).

This is not to say that O’Reilly’s (2011) vision of “government as a platform” is inevitable. Should government become entirely reliant on third-party services, infomediaries themselves may become the new platforms that government plugs into, not the converse. Cloud platforms bring about the risk of being locked into proprietary APIs that limit a government’s ability to outsource other components or substitute the platform itself (Paquette, Jaeger, and Wilson 2010). Infomediaries also are not equally substitutable. Sangiambut and Sieber (2016) found that developers came with differing agendas, particularly those with anti-corruption and transparency mandates built into their missions. As such, government may not easily replace a non-profit advocacy’s open data service with that of a for-profit business.

Government also may experience a highly inelastic demand for services such as open data platforms, particularly when open data is mandated by data directives. Efforts towards government data that is “open by default” (Government of Canada 2012) may pull open data to the centre of mission-critical government functions. Should the price for portal services increase substantially, government may need to pay that cost. High dependency coupled with inflexible demand may induce monopolistic or anti-competitive behaviour by developers seeking to strengthen or maintain market share, especially if they are not easily substituted.

CONCLUSION

Through an exploration of open data app ecosystems, we identified infomediaries and categorized them by type and function. Sources of control over data and data flow are not necessarily between one government and one outsourced app developer but exist in a complex network of mutual interdependencies. We began by examining human actors but found that non-human actors exerted considerable, and sometimes unacknowledged, influence in the ecosystem. With sufficient trust, governments expressed low concern over the risks of government “downsizing” payments to the portal company due to austerity (Sieber and Johnson 2015) or private-sector firms terminating services due to lacklustre business models.

Infomediaries are increasingly crucial in an era of interconnected and cloud-based IT. Compared to traditional software outsourcing, current apps built on open data, crowdsourcing, and Web 2.0 require more complex interdependencies. These arrangements are further supported by existing governance paradigms such as citizen-sourcing and NPM. The more mission critical a service (e.g., an unofficial transit app upon which everyone now relies), the more government should be mindful of the infomediaries and the resultant interdependencies of the ecosystem.

Conflict of Interest

All cities mentioned in this paper are affiliated with a multi-researcher Social Sciences and Humanities Research Council (SSHRC) partnership grant (895-2012-1023), which funded this research. The respondent from Open North and one respondent from the City of Vancouver are the contacts for the grant. However, there was no obligation to participate in this study.

About the Authors

Suthee Sangiambut completed his M.A. in McGill University’s Department of Geography. His graduate research was related to open government, open data, and civic apps, particularly their impacts on citizen-government interactions, government practices, and policy. He continues his work in open data and open government policies in the non-profit sector.
Renée Sieber is an associate professor of geographic information science, jointly appointed between the Department of Geography and the School of Environment at McGill University. She received a Ph.D. in planning and public policy from Rutgers University. Her research interests include public participation geographic information system, community engagement, governance, open data, and civic technology.

Corresponding Address:
Department of Geography
McGill University
Montreal, QC, Canada H3A 0B9
(514) 398-4941
renee.sieber@mcgill.ca

References


Upcoming Conferences

March 19-22, 2018
2018 GIS/CAMA Technologies Conference
Houston, Texas

April 9-13, 2018
URISA GIS Leadership Academy
Columbus, Ohio

July 23-27, 2018
URISA GIS Leadership Academy
Salt Lake City, Utah

October 8-11, 2018
GIS-Pro & CalGIS 2018
Palm Springs, California
GIS-Pro & CalGIS 2018
October 8-11, 2018
Palm Springs, California

Save the Date!