Open Geospatial Data Flows of Civic Apps

Suthee Sangiambut[1], Renee Sieber[2]

1 Geography, McGill University, suthee.sangiambut@mail.mcgill.ca
2 Geography, McGill University, renee.sieber@mcgill.ca

Abstract

The provision of open government (geospatial) data is purported to be an enabler of innovation. Data is flowing from government to citizens in the form of open geospatial data release, which government hopes will promote economic growth and civic engagement. Data is also flowing from the public back to government through crowdsourcing and public consultation applications. However, data does not remain static in its flow to or from government; transformations occur throughout its path that affect structure, content, potential re-use, and assumptions behind data. Using the app as the unit of analysis, five Canadian municipal civic apps representing Government-to-Citizen (G2C) and Citizen-to-Government (C2G) data flows were selected. Through interviews of key respondents in government and developer organizations, we followed the data in its path between government and app developer. Results revealed multiple origins and destinations of data, while data transformations occurred inside and outside of government. Examining processes within open geospatial data flow can reveal its potential and limits in enabling economic growth and civic engagement.

Background and Relevance

Open data is data that "can be freely used, modified, and shared by anyone for any purpose" (Open Knowledge International, 2012), and recent accounting of open data catalogues in Canada strongly indicate that open data is predominantly spatial and GIS-compatible (Baculi, 2014; Baculi & Rinner, 2014). Whereas the flow of open data through open data catalogues appears to be unidirectional from government-to-citizen (G2C), government data production already includes third party sources such as the case of volunteered geographic information (VGI) (Goodchild, 2007) or even older channels such as 311 service request hotlines. Recently, Statistics Canada has implemented its own crowdsourcing initiative using the VGI platform, OpenStreetMap (Canada, 2016). This inclusion of VGI into the open data dialogue suggests a shift towards open geospatial data being sourced internally and externally of government and a bidirectionality of data flows.

Arguably the locus of interaction with open data is not likely to be with datasets or data portals themselves, but rather through software applications (apps) that repackage and provide an easy-to-use interface to view or collect data. Infomediaries may transform open data into more communicable and accessible forms in apps for the purpose of increasing civic participation (Sangiambut & Sieber 2016).

These civic apps may be developed by government or outsourced to the private sector, non-profits, or general public. For example, the City of Toronto does not develop its own 311 service request app and instead recommends three third party apps to the public while relying on the Open311 API (Application Programming Interface) as an interface;

also outside their control[1]. Sieber and Johnson (2015) say one of the big problems is that government often conceptualizes open data as essentially "throwing it over the wall", publish the data and you're done with it. Even in the act of releasing data, government may lose the ability to control or promote outcomes of data use (Janssen et. al., 2012). Davies and Frank (2013), in their exploration of 'raw' open data, note that open data are "constructed data, potentially brought together from many flows of data inside government, and that much of what goes into an open datasets construction remains opaque in current practices" (ibid., p. 75). Their findings on data creation, data release, and data re-use revealed a decision over data formatting or release could be implicit or happenstance. Their data also suggested that data formatting and structure were not tailored to anticipated data re-use cases (ibid., p. 77). A past paper (Sangiambut & Sieber, 2016) examined the nodes of through which data flow, in the form of infomediaries, through the rhetoric of government outsourcing and Actor Network Theory. For this paper, our goal was to trace the flows between nodes as distinct elements unto themselves. Prior research explored civic apps as a network of power relations among actors (Sangiambut & Sieber, 2016). In this paper, we focus on the flow of data itself and the transformations it undergoes from origin to destination. This focus on data flow is different from the concept of a network of actors, as it can help us understand resulting outputs and outcomes of open data, instead of the power relations between infomediaries and governments. Open data can be both a process and an output unto itself. Data can have multiple origins, especially if it represents an aggregation of information or is transformed into a new data product or becomes an input into another product. It is not always obvious who provides data and what sector they are from; we reveal that multiple sectors can be aggregated into a single dataset.

Data that undergoes too much transformation or is re-used as an input into another product, may no longer reflect the intent of its originator and therefore may find its use as a catalyst for innovation and civic engagement diminished.

Methods and Data

Five Canadian municipal civic open data app were selected based on a typology of apps (Sangiambut & Sieber, 2016), which outlined ideal types based on directionality of data flow (G2C, C2G) and the role of the app user (economic or political). To explore the data flow in these apps, a 'follow the data' approach was adapted from Davies and Frank (Davies & Frank, 2013). In their study, Davies and Frank (ibid.) tracked a single dataset back from its public face to its source to understand its release and reuse. Their method informed the framing of interview questions and the recruiting of respondents. In interviews respondents were asked to describe the history and development of the app from its institutional perspective (government or developer), its data structure and reasoning behind data transformations. Respondents also were asked for their perceptions on civic engagement enabled by the app, and their views of their relationship with their developer or government counterparts.

---

[1] http://www1.toronto.ca/wps/portal/contentonly?vgnextoid=bc6bfbfa98491410VgnVCM10000071d60f89RCRD

Table 1. lists the apps, the data content, the directionality of data flow, as well as the government and developers involved in the development of the app. Our method was inspired by Latour's Actor-Network Theory (Latour, 1987) in the identification of non-human actors as well as human actors through which the data flowed. Non-human actors, like software, act as "inscription devices", which can "transform pieces of matter into written documents" (Latour & Woolgar, 1986, p. 51). Such actors, including software, have the agency to exert their own influence on data flow.

Table 1. List of Apps and Associated Institutions

App	Citizen Dashboard	Citizen Budget	Ottawa Transit	Toronto Cycling App	VanConnect
City	Edmonton	Montreal	Ottawa	Toronto	Vancouver
Developer	Socrata	Open North	3lywa Solutions	Brisk Synergies	PublicStuff
Data flow	G2C	C2G	G2C	C2G	C2G
Data content	City administration performance measures	Budget consultation	OC Transpo bus data	User cycling routes	311 service requests

Results

In this section, we cover the results of interviews exploring data flow with government and developer respondents, some of which confirmed findings from Davies and Frank (2013). Results showed that both G2C and C2G open data flows experience transformations outside of government control, such as through third party developers. We were surprised to find that data had multiple origins and destinations inside government; data production and release could therefore be backed by multiple intentions and assumptions. Moreover, the initial assumptions a data originator has may not survive successive data transformation down the line. Findings were arranged according to four categories: origins, transformations, release, and re-use (Table 2.). These categories were based on Davies and Frank's (2013) framing, which we found helpful to demonstrate data flow's characteristics as both process and output. For this reason, we do not focus on a singular type of transformation or case study.

For brevity, we highlight findings from each of these categories taken from across all cases, and group them according to the directionality of data flow (G2C/C2G).

Data flow could be serial whereby data must pass through a series of gates in succession, such as the case of Citizen Budget (Figure 1.), or radial where data could end up in a number of different destinations such as Toronto Cycling App (Figure 3.).

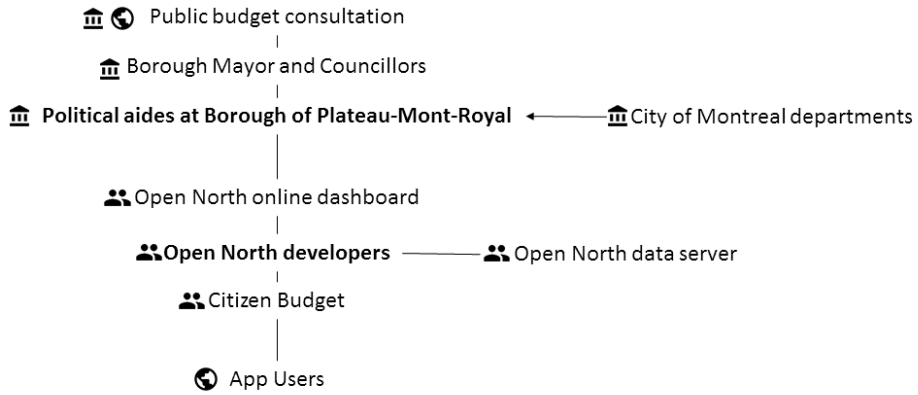

Figure 1. Citizen Budget app ecosystem (Sangiambut & Sieber, 2016)

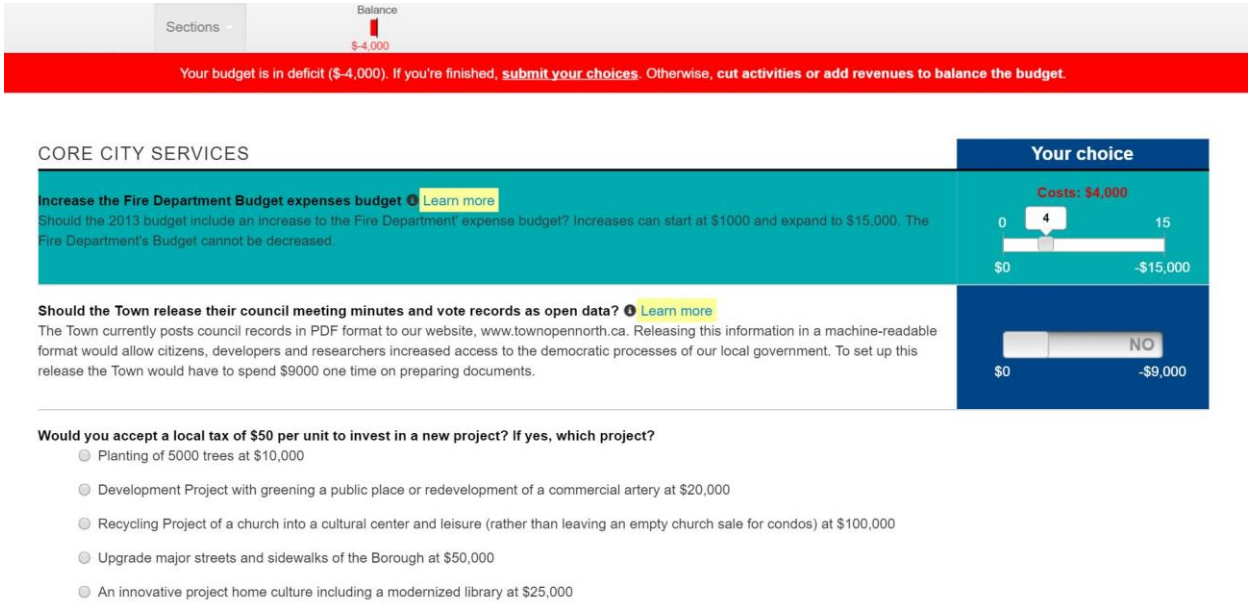

Figure 2. Citizen Budget interface (Sangiambut & Sieber, 2016, p. 147)

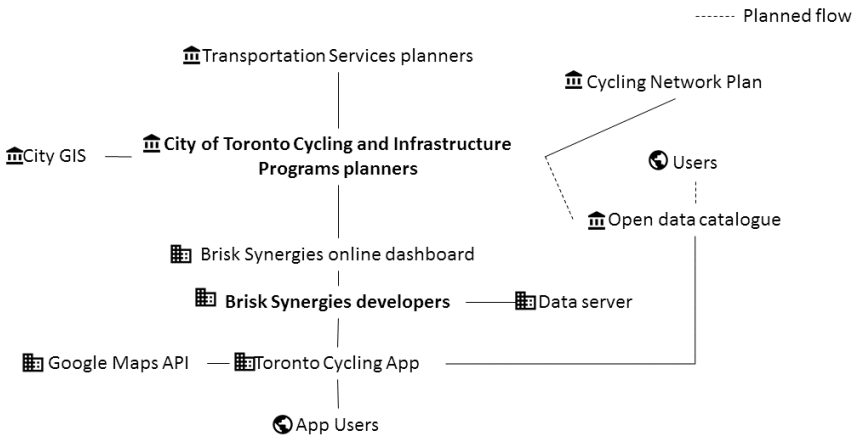

Figure 3. Toronto Cycling App ecosystem (Sangiambut & Sieber, 2016)

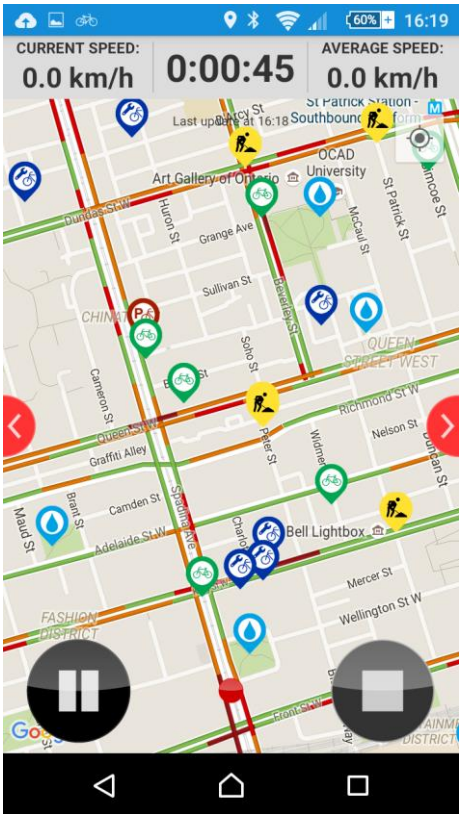

Figure 4. Toronto Cycling App interface (Sangiambut & Sieber, 2016, p. 147)

Table 2. Characteristics of Data Flow

	Directionality of Data Flow	
	G2C	**C2G**

Origins	Multiple government departments and civil servants Singular software platforms	Broad public: targeted and self-selected
Transformations	Data hosting, structuration, formatting	Geocoding
Release	Data broadcast to the public	Re-release can require aggregation
Re-use	Assumed broad, unspecified re-use Informal tracking of re-use	Specific re-use cases in government

Data can originate from multiple sources. For C2G flows, data originated from app users, either filling in a form (e.g., budget simulation in Citizen Budget) or through the app's collection of data from smartphone sensors (e.g., GPS data in Toronto Cycling App). Because data contributions all had some level of geospatial identification (e.g., latitude and longitude, street address, postal code), data flow could be attributed (to some level of certainty) as coming from specific geographic areas, bringing in an element of locality to data origination.

For G2C, data could originate from multiple branches of municipal government; Citizen Dashboard's data is provided by data stewards from all City of Edmonton branches. Data could also originate from single entities within government. OC Transpo's bus schedule data is produced through a Computer Aided Dispatch/Automated Vehicle Location (CAD/AVL) software platform that is central to the entire agency's operations.

Data underwent such processes as structuration, formatting, and geocoding. Influence over these processes, and by extension the content of data, was evident in both government and developers. Data flow could also be controlled in terms of velocity such as through API token management, which can establish a download quota. Individuals or organizations with the decision-making power can also control data flow.

Both virtual and physical stores for data are found in open data catalogues and the physical servers they reside in. In the case of Ottawa Transit, an SQL database was used to store data, which required a transformation of GTFS data into a relational database that could be queried. This suggests that open data flow is not simply a linear path from government to users. The introduction of outsourced developers can complicate data flow beyond what is visible.

In terms of formatting, G2C data required outputting into specific data formats and structure that could be publicly consumed, such as the General Transit Feed Specification (GTFS). Such transformations can be necessitated by industry standards – GTFS itself has become the de facto standard for displaying transit data, even though proprietary CAD/AVL systems such as Hastus work on completely different data structures. APIs can also transform data structure. Socrata's Open Data API was used to upload and update data in the City of Edmonton's open data catalogue (Figure 5.). The API influences data structure by requiring data to be tabulated, and does not have capabilities to properly display qualitative data. This resulted in the City's credit rating performance simply not being visualized[2]. Restrictions on data structure could therefore

[2] https://dashboard.edmonton.ca/en/stat/goals/yae9-vbqu/j3qs-ebqi/vwzq-e2u2

affect final data visualization in an app. For more information on non-geographic transformations, see Sangiambut and Sieber (2016).

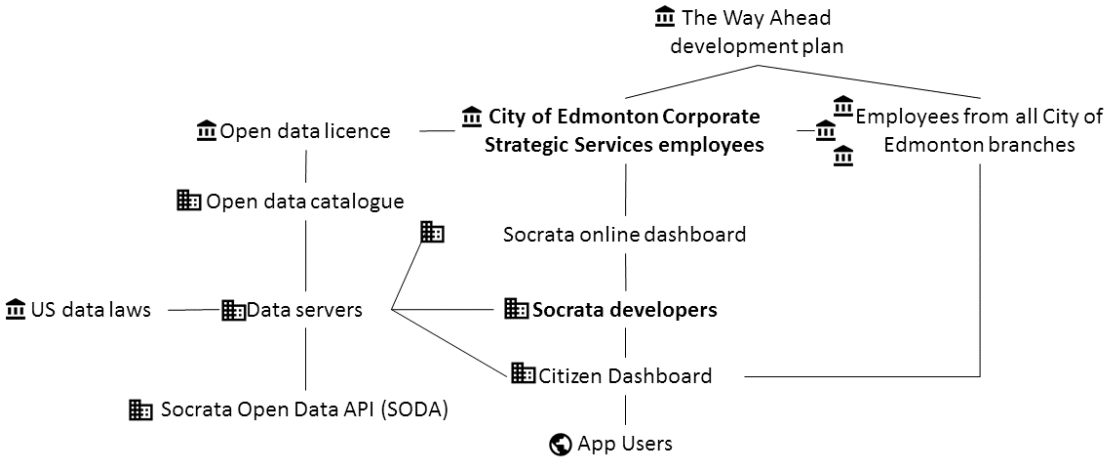

Figure 5. Citizen Dashboard app ecosystem (Sangiambut & Sieber, 2016)

Geocoding (and reverse geocoding) and mapping were under-emphasized transformations. The Google Maps API could be used for data visualization, geocoding, and even network routing calculations. This occurred in both G2C (Ottawa Transit inputs stop location into a Google Maps API query to calculate routes for trips) and C2G (VanConnect performs geocoding of service requests via Google Maps API) data flows.

Approaches to data release were also noted. G2C flow such as the case of Ottawa Transit and Citizen Dashboard could be considered a 'broadcast' method, whereby data is simply thrown 'over the wall', leaving it the responsibility of citizens to view data. Data release in C2G flow depended on the need to anonymize and aggregate data. For 311 service requests, government respondents felt the need to anonymize service requesters by removing names and reducing the accuracy of request locations to street names and blocks. However, data release within government for 311 service requests is tailored for civil servants to be as accurate as possible, with all the details of the service request, including GPS coordinates, forwarded to the relevant infrastructure maintainer.

Finally, we noted the potential scenarios for data re-use. Government respondents for G2C flows were optimistic over the potential for data re-use and were willing to speculate on potential scenarios, but felt it appropriate to leave re-use up to citizens. On the other hand, C2G flows were all designed for very specific government use cases (such as a cycling network plan) and specific government units such as the City of Toronto Cycling Infrastructure and Programs unit.

Conclusions

By following the flow of data on its journey between developer and government, we have shown that open data, whether it is G2C or C2G, continues to be transformed after leaving its origin on its way to its destination. This can occur both inside and outside of the government institution and jurisdiction. The ability to re-use data is crucial in sustaining open geospatial data initiatives. Awareness of such transformations is important in understanding whether intended outcomes of open data can be fully realized and how to make government data resilient to influences of the private sector. When data transformations occur or originate from outside government, cities may find themselves limited in their ability to promote data re-use.

References

Baculi, E. (2014). Content and Growth of Municipal Open Data Catalogues. Retrieved from http://www.gogeomatics.ca/magazine/content-and-growth-of-municipal-open-data-catalogues.htm

Baculi, E., & Rinner, C. (2014). *A Geographic Content Analysis of Municipal Open Data Catalogues across Canada*. *Canadian Cartographic Association*.

Canada, S. (2016). Crowdsourcing. Retrieved from http://www.statcan.gc.ca/eng/crowdsourcing

Davies, T., & Frank, M. (2013). "There"s no such thing as raw data'. In *Proceedings of the 5th Annual ACM Web Science Conference on - WebSci '13* (pp. 75–78). New York, New York, USA: ACM Press. http://doi.org/10.1145/2464464.2464472

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal*, *69*(4), 211–221. http://doi.org/10.1007/s10708-007-9111-y

Janssen, M., Charalabidis, Y., & Zuiderwijk, A. (2012). Benefits, Adoption Barriers and Myths of Open Data and Open Government. *Information Systems Management*, *29*(4), 258–268. http://doi.org/10.1080/10580530.2012.716740

Latour, B. (1987). *Science in Action: How to Follow Scientists and Engineers Through Society*. Cambridge, Massachusetts: Harvard University Press.

Latour, B., & Woolgar, S. (1986). *Laboratory Life: The Construction of Scientific Facts* (2nd ed.). Princeton, New Jersey: Princeton University Press.

Open Knowledge International. (2012). OpenDefinition. Retrieved from http://opendefinition.org/

Sangiambut, S., & Sieber, R. (2016). The V in VGI: Citizens or Civic Data Sources. *Urban Planning*, *1*(2), 141–154. Retrieved from http://dx.doi.org/10.17645/up.v1i2.644

Sangiambut, S., & Sieber, R. (2016) *The Civic Open Data and Crowdsourcing App Ecosystem: Actors, Materials, and Interventions*. Manuscript submitted for publication.

Sieber, R. E., & Johnson, P. A. (2015). Civic open data at a crossroads: Dominant models and current challenges. *Government Information Quarterly*, *32*(3), 308–315. http://doi.org/10.1016/j.giq.2015.05.003