Open Data in Quebec's Public Administration:

Uses, obstacles and potential solutions

By

Christian Boudreau, Ph.D., professor, ENAP
Daniel J. Caron, Ph.D., professor, ENAP

École nationale
d'administration
publique

NetGouv Open Data
Measure of the contribution and evolution of open government in Quebec

This report is part of the *NetGouv Open Data* project whose purpose is to develop, implement and disseminate the results of a new measure of the usage of government data in Quebec and to assess how the measure promotes economic growth. The project aims to measure adherence to the new open government data approach within various ministries, government agencies, and municipalities. The report also looks into internal collaboration with other ministries and agencies, as well as open data planning and deployment.

Project Team – CEFRIO
Claire Bourget, Senior Director of marketing research
David Huynh Quan Suu, Project Authority in marketing research
Guillaume Ducharme, Director of Communications
Julia Gaudreault-Perron, Project Manager
Annie Lavoie, Communications Advisor

Scientific Team
Christian Boudreau, Ph.D., professor, École nationale d'administration publique
Daniel J. Caran, Ph.D., professor, École nationale d'administration publique

TABLE OF CONTENTS

INTRODUCTION

The public service is undoubtedly one of the largest producers and custodians of data of public interest. Its data covers a range of areas of activity with which it is engaged, such as transportation, health, public safety, the economy, the environment, culture, education, agriculture and a host of other sectors. By liberating its data and making it easily accessible and usable, the public service seeks to stimulate the economy through the development of mobile applications while at the same time improving its low-cost service offerings and promoting transparency in its institutions. Open Data would bring multiple benefits (e.g.: economic, democratic and administrative) thanks to its (re)use[1] by various societal actors (e.g.: citizens, journalists, businesses, NPOs, public service). By "open data" we mean digital data, often granular[2] and free, that public organizations make accessible in formats and conditions that ensure its unfettered use, re-use and redistribution by any and all.[3]

Despite the buzz surrounding it over the past few years internationally, Open Data remains, to a large extent, an emerging phenomenon. Beyond government statements on the matter, public services face a number of obstacles to data liberation, namely the production, formatting and dissemination of open data. If most large Western cities and liberal democracies have set up open data portals, several of these websites contain datasets whose quality, relevance and degree of openness are sometimes lacking. Indeed, it is not uncommon to find data that is either obsolete, poorly documented, or in formats that are difficult to reuse (e.g.: PDF).

Inspired by the Obama administration in 2008, the Quebec government instructed the main provincial ministries and agencies (hereinafter 'agencies') to liberate at least three datasets by 2012 in order to file them in Quebec's new open data portal.[4] While the majority of these agencies heeded the call, few datasets have been added to the portal since its creation. Here as elsewhere, data liberation does not appear to be a movement that is intuitive for the public service. Although a number of representatives, at least those whom we met during the course of this study, have a fairly good idea what open data is, the mechanics of open data and its benefits appear tenuous: How do we go about liberating data? Where do you start? Which data should be liberated? Do the benefits exceed the associated costs? Is it worth the effort? These were some of the most commonly raised questions by these representatives.

---

[1] The re-use and use of open data are two separate activities: the first refers to the original use of open data in order to create applications and services, while the second involves the visualization and querying of the data. Re-use can lead to the development of visualisation and querying tools in order to simplify the use of open data.

[2] Granularity is essential to the use of datasets. Moreover, aggregation by governments in their anonymization processes has rendered the re-use of several published datasets moot.

[3] Definition adapted from Open Data Charter.

[4] More recently, the Quebec government launched, on April 5, 2016, the *Données Québec* portal, which integrates open data from the previous government portal and portals of the following five Quebec cities: Montreal, Quebec City, Sherbrooke, Laval and Gatineau. Upon its creation, this provincial and municipal portal had 770 datasets.

Based on about twenty interviews conducted between May and October 2015 with representatives (hereinafter 'participants')[5] from a dozen Quebec government provincial agencies,[6] this study examines Open Data from three vantage points. First, it looks into potential uses of open data, as expressed by the participants of the study. Second, it addresses organizational conditions that tend to be barriers to data liberation within Quebec's public administration. Third, it proposes potential solutions in order to facilitate the liberation of open data in the public service.

---

[5] Participants' comments are represented as citations in the text.

[6] The following agencies were approached in the course of this study: the Treasury Board Secretariat, Société de l'assurance automobile du Québec, the Ministère de la Santé et Services sociaux, the Ministère de l'Économie, de l'Innovation et des Exportations (now the Ministère de l'Économie, de la Science et de l'Innovation), Revenu Québec, the Ministère de Culture et des Communications, the Ministère de la Sécurité publique, SOQUIJ, and the Institut de la Statistique du Québec.

1. POTENTIAL APPLICATIONS RELATED TO THE (RE)USE OF OPEN DATA

1.1 DEVELOPMENT OF APPLICATIONS AND SERVICES

Although the benefits of open data were not always clear to the participants, they recognize a certain economic, even commercial value of developing applications, particularly for mobile devices. As the participants pointed out, data (or datasets) may appear of little interest when viewed in isolation, but these same data (or datasets) can lead to the creation of new knowledge and new products or services when combined in an original manner.

[translation]

The real benefit of open data is the power to amalgamate and create new knowledge and applications … We created something that did not exist before.

If the development potential can vary from one open dataset to another, there appears to be a consensus on geospatial (or geomatic) data among participants. Geospatial data help localize, that is to say, locate geographically various relevant pieces of information by means of a mapping interface. For one respondent, the potentialities of geospatial data are cross-sectional, in the sense of where geographic positioning enables the layering of data from several fields or areas of activity of the public service.

[translation]

It might be tempting to view geomatics as a speciality. However, in the context of open data, it's almost a normality, no matter which field the data comes from.

Thus, geospatial data may be of interest to individuals and organizations who will have an opportunity to superimpose different layers of data onto a map based on the needs of the moment.

[translation]

It has become normal for citizens to have access to data that are useful to them locally or wherever they may find themselves. The location of daycare centres or bus routes may interest them … or the crime statistics for certain neighbourhoods, from Sûreté du Québec data … If you liberate data on routes, addresses, buildings, bike paths, snowmobile trails, walking trails, lakes … some people will surely create interesting mobile applications.

Geospatial data may also be useful for ensuring public safety after natural disasters (e.g.: flooding in the Richelieu valley) or humanitarian tragedies (e.g.: Lac-Mégantic). Public authorities are increasingly reliant on such data in crisis situations in order to intervene effectively on the ground. For example, the Ministère

de la Sécurité publique [translation] "provides 911 centres with a mapping interface that allows them to display several layers of information … in order to be able to locate individuals who are in difficulty."

A number of participants noted that through the development of applications, Open Data can contribute to economic development and improvements in the way public services are provided, at a low cost to the government; one need only consider the many applications developed in the areas of public transit, meteorology, agriculture, employment, the economy, health, justice, as well as public safety and emergency planning. For some participants, it is clear that the government cannot rival the private sector's creativity, agility and ability to take risks in developing applications.

> [translation]
>
> *When the government develops an application, it does so cautiously … development costs are much higher in a government context … than in the private sector ... If a business makes an error in the application … no one will get upset. But when a department makes one, it will make the front pages … As soon as the department develops something, you have big government getting involved.*

Government initiatives, such as the creation of the open data portal and the organizing of *hackathons*,[7] are aimed at stimulating the development of all sorts of applications, of sometimes unexpected uses, by various economic and other societal actors (e.g.: NPOs), according to the respondents. Through these initiatives government seeks to solicit the expertise of creative companies, beyond its borders, to contribute to improving services without being too costly.

1.2 GOVERNMENT TRANSPARENCY AND CIVIL OVERSIGHT

Data liberation is also associated with benefits to democracies. Indeed, it can be a mechanism for more government transparency and accountability by highlighting public service performance and, more specifically, the quality of its services, the efficiency of its resources and the effectiveness of its programs. In a context of budget restrictions that call for tight resource management, agencies are faced with little choice but to be increasingly transparent. By liberating data on their operations, programs and services, agencies "expose themselves" to increased oversight by various societal actors who are able to follow their activities closely and analyze their decisions. Open Data is often depicted as a way of diminishing the risk of corruption and collusion in that it allows for better monitoring of the public procurement process, in particular, the management of government contracts with private companies. Data liberation of the système électronique d'appel d'offres du government du Québec (SÉAO)[8] is part of this strategy of transparency and monitoring of public procurement in Quebec, according to participants. With data from the SÉAO,

---

[7] Event normally held over the course of several days in which users and developers build applications, in particular by using open datasets.

[8] Agencies are required to publish certain information related to the number of contracts they award that are worth $25,000 or more.

[translation]

A company is able to see the price of the contracts that were awarded in its region … It can then adjust its contractual strategies in kind … or report suspicious circumstances … especially when the same bidders are always awarded contracts.

Several participants shared the view that journalists are among the actors who were most likely to use Open Data to keep track of the government's activities, judging by their interest in public affairs and by the large number of access to information requests made to agencies. [translation] "We receive about a thousand media requests every year," stated one agency's communications lead. When an individual is identifiable by the data [translation] "we will suggest that they file an access to information request," he added.

Researchers were also perceived by participants as being significant potential users of open data. Free access to large volumes of raw data can only make it easier for researchers to do their work. A number of participants noted that their organization already calls upon researchers to study issues or themes that concern them. Depending on the breadth and quality of the open data, researchers may be able to analyze the way programs and public services function so as to evaluate their performance.

1.3 ADVANTAGES FOR AGENCIES

It can be taxing for agencies that have a lot of data to respond to access to information requests from journalists, researchers, and other users. It was noted that the volume and complexity of requests tends to increase with the computerization of files. [translation] "Certain requests can be burdensome, very burdensome to process," noted the statistical data lead of one agency. Open Data was viewed by participants as a way to reduce requests for information by enabling individuals to [translation] "do their own searches … which means there is one less professional … spending time responding to the request." In order to simplify the use of open data and promote users' autonomy, participants suggested implementing simple visualisation and query tools: [translation] "We could build a nice little pivot table … users would only need to choose the variables they want."

The (re)use of liberated data is not only of interest to non-governmental actors. What an agency produces in data may also be of interest to others and may help enlighten their decision-making process.

[translation]

For example, the deprivation index in schools, which is a dataset published by the Ministère de l'Éducation on a yearly basis … can help the Ministère de la Famille identify where to set up daycare centres.

Agencies may also show an interest in Open Data produced by other government bodies when they are developing or reviewing their own programs. By examining probative data on government initiatives in

other countries, it becomes possible for an agency to predict the effects of certain program changes on its clientele without having to experiment. For example:

[translation]

The United States has an open database they call FARS … on fatal road accidents … That dataset enabled us to do our own analysis … and to measure the impact of increasing speed limits on highways.

Open Data also provide governments with a way to make comparisons between themselves, particularly when the data are well documented in terms of methodology and metadata. [9] [translation] "Other provinces may want to compare themselves to us, perhaps even other countries." We were told that comparisons and benchmarking becomes even more accessible when the data selected meet international standards: [translation] "We now have international agencies such as the Open Government Partnership issuing standards … we are able to benchmark."

In short, the potential uses and benefits of open data for its users can be many and varied. Given the economic, administrative, democratic and scientific advantages it provides, the data liberation movement [translation] "can only increase," according to one of the participants. [translation] "It's the beginning of something significant," added another. However, data liberation and the government transparency that comes with it pose significant challenges for an organization that governments' messaging tends to overlook, preferring to focus on the benefits. The next section will examine, again from the perspective of Quebec's public administration, the organizational barriers to open data liberation.

---

[9] A metadata unit defines the characteristics or attributes of a dataset. It specifies, among other things, the contents of the dataset, its format, when it was updated and its author(s) or contributor(s).

2. ORGANIZATIONAL BARRIERS TO DATA LIBERATION

2.1 OTHER PRIORITIES

Aside from the creation of the portal in 2012, the efforts at data liberation made by a number of agencies appear to have slowed down. Most of the participants in fact indicated that Open Data arrived near the bottom of their organizations' list of priorities: [translation] "For the first two and a half years … it was really not a priority,'' stated one participant. Two background factors tend to explain this situation. First, there was a change in government a few months after the open data portal was launched in the summer of 2012, which led to a shakeup of the government`s priorities on the matter.

[translation]

When it was launched in June 2012, the will of the Liberal government at the time and of Madame Courchesne was very clear … However, in the years that followed … we did not get that same sense from the Parti Québécois government.

Furthermore, several participants maintained that their organization lacked the necessary resources to start projects that require new investments and whose benefits are not immediate, such as with data liberation, given the context of budgetary restrictions and the increased workload it would entail. [translation] ''We are more in a staff-cutting mode than an expansion mode … we don't have the staff to support those kinds of things.'' The organization's priorities lie elsewhere, we were told, primarily in maintaining and consolidating current operations. Senior management of the agencies we met with placed very little pressure on those individuals who were designated as open data stewards, commonly referred to as respondents.

[translation]

Aside from the year it was launched, no one told me: go for it … My mandate was to collaborate, to publish data, but not to make it priority Number One. We had much bigger fish to fry.

The data liberation movement is the result of limited individual or group initiatives rather than a structured organizational strategy, according to most of the participants. [translation] ''There is a willingness, but it is more individual or group-based,'' added one participant; ''we are pretty much into the initiatives and the professional freedom that take care of that,'' added another. Although most open data stewards with whom we met had a committee[10] to support them in their work, the data liberation project in their organizations and the conduct of such activities fell largely to them. In this trend towards individual

---

[10] Committees are usually made up of representatives from various sectors in the organization: communications, technology, business lines, access to information, management, etc.

initiatives, one open data steward had to develop his own manual process and search the Internet to find the tools to liberate his own organization's geospatial datasets.

[translation]

It does not happen automatically … the last time, updating seven files took me three days … there were times when I had to go home, because here I didn't have Google Earth … when you are a government, you need a license … I would do it at home in the evening … I drafted a procedure that was eight pages long in Word.

Open data stewards perceived their work primarily as an evangelization exercise that they undertake often alone, without any significant support from senior management. For a number of them, this awareness-raising exercise begins from square one every time they approach a new member of their organization to introduce them to open data or to have them participate in any way in freeing up data.

2.2 POTENTIALLY DEMANDING OPERATIONS

Freeing up data can be a demanding process, particularly when it means choosing datasets with a potential for (re)use. This presupposes a solid knowledge of the information the organization has. An inventory of data is a prerequisite for their liberation; [translation] "it's the first thing that must be done before determining the data to be freed up." Once the inventory has been completed, choosing the data remains a complicated operation due to the numerous discussions and negotiations this may generate internally. For many participants, the selection process is by far the most sensitive step [translation] "the most sensitive … if you spend $100,000, you will spend $90,000 debating and $10,000 on technologies to publish it." This step is all the more contentious in that the data to be freed are used to demonstrate accountability, stated one participant who has experience with freeing up this type of data.

[translation]

Because these are accountability data … and not statistical data on the aging population … it takes an unbelievable amount of time … it took numerous high level discussions … then, given that the data are subject to later interpretation … it's negotiated … with Litigation.

The preparation and updating of some datasets can also be a highly complex operation when, among other things, Open Data cover all government operations, such as public expenditures, and that these data require frequent updating.

[translation]

The government's expenditure budget is five volumes … It is updated on a daily basis by means of the Treasury Board. Envelopes, budgets, and program elements are adjusted … through about 1,800 decisions per year … putting that online is a colossal

effort … especially since the data are not in an easily useable format to begin with … they're in PDF.

Data liberation can also elicit various questions from users who expect a response from agencies and, therefore, a certain level of customer support. For example, users may detect errors in datasets and notify the custodial agency. [translation] "If the dataset … contains errors–one should expect calls, which take time." Users are also likely to call the agency that freed up the data when they notice variations between the open data and other comparable data to which they have access. The same thing can occur if the data are produced by different bodies in varying circumstances.

[translation]

It's not just putting the data online, then the phone rings! 'Why is it written … upon studying the credits ... in your annual management report … when you were asked about it in the meeting that's not what you said' ... Major efforts are required for post publication support ... In certain cases this leads to corrective measures ... sometimes you need to go all the way back to the source.

Lastly, some participants fear seeing users mistakenly interpreting open data as a result of limited methodological skills, particularly in the interpretation of statistics. In their view, freeing up data will lead to their organization, specifically its research and statistics production service, having to respond to various questions from users when there are discrepancies or misunderstandings with respect to the interpretation of the data.

[translation]

If people become more autonomous ... they may arrive at conclusions that are at odds with ours ... Not everyone is comfortable doing that ... This will require us to explain the methodology and interpretation: 'How is it that you extract one type of data but then when I do it I don't get the same result at all!'

2.3 TECHNOLOGICAL OBSOLESCENCE AND DILUTION

In the agencies that were studied, as in Quebec's public administration generally, there are many aging technological infrastructures. Some information systems, which have been in existence for over thirty years, have not been modernized (or have been only very slightly modernized) since their creation, including a number of mission systems that generate agencies' programs and public services. Aside from the fact that [translation] "many of these systems have a cycle of obsolescence that is fast approaching," these same systems were not designed for easy data extraction: [translation] "Finding the data, preparing it to become freed up ... can take a lot of time." Moreover, staff who are able to use the aging systems are increasingly rare:

[translation]

It takes people with almost archeological skills ... Young people who came to work there … would leave after a month.

Obsolescence of systems is not the only technological challenge facing managers when data are freed up. They also have to deal with various different information systems and technology platforms, another technological aspect that comes with its own difficulties.[11]

[translation]

One of my datasets is on ORACLE, the other on Microsoft ... so making these data public takes me time ... technological dilution is very much a reality of today.

The dilution of technology is, among other things, related to the fact that public agencies have until now tended to develop information systems specifically tailored to their line of work, we were told. For example, one participant noted [translation] "22 different technological environments" in his organization that were, for the most part, not integrated. This multitude of siloed systems makes it more complicated not only to extract data, but also to store it: [translation] "There is no architect who has an overall vision of all the department's data." The quality of the data is also affected, particularly with regard to the uniqueness of information. This is the case with some of the administrative data needed for managing programs and services, such as the addresses of businesses.

[translation]

We have a lot of businesses' addresses ... there's a lot of duplication ... it needs cleaning up ... Our different services sometimes manage the same variables... There are exorbitant costs involved in matching and integrating everything.

2.4 INCONSISTENT DATA QUALITY

Most of the participants were reluctant to free up data when such data was not of high quality. Admittedly [translation] "several management operations, such as those conducted in budget planning and accounting, are still done using Excel," added one participant. Updating data is one of the quality criteria cited that tends to dampen enthusiasm for data liberation. This criteria appears particularly problematic for high-volume transaction data, given the frequency and extent of the updates. Delays in the data entry

---

[11] However, one participant indicated that [translation] "there are technologies that allow for data in systems (even old ones) to be read without having to modify them."

process can give rise to significant discrepancies between the moment of the transaction and when the data is freed up, as the following comments attest:

[translation]

In the department ... the updating of data can be debatable at times ... Take the example of the financial assistance we process. There are some people in the department ... who don't enter the data [as soon as they receive it] into the mission system ... they enter it at the end ... these data do not reflect reality ... it will take a synchronisation between the transaction and the data we will free up.

Updating can also be problematic in that one dataset may be fed by various contributors from outside the receiving organization. Contributions from several partners can also lead to input errors, thereby affecting the reliability of the data.

When an agency's data [translation] "comes from outside, it [its reliability] is more complicated to verify and it's only later on that you discover it." Often, it is only once the error had been pointed out by users that the agency is able to make the necessary corrections.
For example,

[translation]

… in the case of the SÉAO ... it is people from various municipalities and elsewhere who manually enter the data ... There is nobody to validate the thousands of data items that are entered.

For another participant, the fact that State representatives have to manually enter many administrative data points during the course of incidents makes them prone to errors:

[translation]

Police incident reports ... are entered manually, following the incident ... on an excessively dense form ... at times, my people have found inconsistencies.

In addition, in the production of statistics and performance indicators on their services and clients, agencies may respectively use different methodologies that may confuse users who consult and are faced with these different sources.[12] The uniqueness of information can also be an issue for organization. In this regard, one participant noted that his department had used four different methodologies to calculate delays in access to various services. Lastly, certain administrative databanks managed centrally by the

---

[12] It was this desire for data standardization that led to the creation of the Banque de données des statistiques officielles (BDSO) and which was the fruit of "a collaborative effort involving departments and agencies under the coordination of the Institut de la Statistique du Québec" (BDSO website).

State, staffed by public servants, would benefit from being cleaned up, documented and standardized before being freed up, according to one participant.

[translation]

If we want to go further with a database such as that of public service staffing, there will be a colossal cleanup job to do.

However, it must be recognized that most service agencies (RAMQ, SAAQ, Revenu, RRQ, CSST, etc.) have validation mechanisms at their disposal to ensure the quality of the data they collect in their transactions with clients: [translation] "Our data are filtered, analyzed for any inconsistencies, and compared from one year to the next," noted a representative of one of these agencies.

Moreover, some participants opined that the State need not necessarily wait until its data of public interest is of near-perfect quality (e.g.: exact, exhaustive and up to date) prior to publication, *a fortiori* when the liberation of such data presents no risk to the public.

[translation]

For example, if I want to put in data on bike paths in Quebec, why wait until I have all of the data before publishing it ... Inputting 75% is not too bad ... especially when ... there are no negative effects from the fact that the data is not 100% complete.

For these participants, the fact of making the data available as soon as possible may be a way to improve its quality through feedback from users. Thus, the quality of a dataset tends to increase with use. This kind of crowdsourcing of users has had beneficial effects on the quality of an important government databank:

[translation]

The best example is ... the Adresses Québec databank ... We started with something that was, say, 65% complete ... Users, along with people from Adresses Québec, worked to improve the product ... We ended up with a government of Quebec databank that contains 3,400,000 addresses ... Almost every address ... In short, you don't need to wait until it is exhaustive, otherwise you won't get very far.

2.5 SENSITIVITY OF INFORMATION AND CULTURE OF SECRECY

Under various different laws and regulations, a certain amount of government data is confidential in nature, particularly personal information, which spurs agencies to put in place security measures to protect the identity of their clients (e.g.: citizens, professionals or businesses). When freeing up transactional data between the State and citizens, the right to that information quickly runs into the

privacy right of members of the public. With such data, agencies see themselves more as guardians, rather than providers, of information.[13]

[translation]

I'm more about protecting the data … It's more about controlling the information … That kind of information must be protected at all costs.

Even when confidential data is kept anonymous, participants remained prudent about its liberation, citing the possibility of re-identification through cross-referencing the increasing amount of open data of members of the public. Confidentiality can also extend to communities when liberated data may cause harm through the effects of social stigmatisation; to which one participant added:

[translation]

Data can stigmatise … certain groups of people … such as information about the social/material deprivation index of a given area.

The principle of confidentiality is not limited to the protection of sensitive information on individuals, communities or businesses to whom the State provides services or with whom it does business. Data on government activities deemed strategic can also be classified as confidential, as provided under the Access Act.

[translation]

Decisions made by the Treasury Board are confidential for a 25-year period. So are the analyses that preceded the decision … I won't disseminate raw data that would effectively reveal my negotiating strategy … Nor will I reveal my datasets … with regard to the application of a budget policy that will affect the economy.

Moreover, participants noted a reluctance on the part of organizations to release data on their performance. For them, the liberation of this type of data is likely to be subject to an intensive process of validation and authorization by their administrative and political authorities.[14] The inclination of agencies to control the message and image leads them to be prudent, even secretive, especially when it comes to liberating data (e.g.: performance indicators) that puts accountability in the public eye.

---

[13] In the event that operational data contain personal information, its transfer from organization to another must be subject to an official agreement, generally approved by the Commission d'accès à l'information pursuant to section 70 of the *Act respecting access to documents held by public bodies and the Protection of personal information* (also referred to as the Access Act).

[14] [translation] "When we release data, we release it to politicians first … if a journalist puts it out, at least we have given the politician advance notice."

[translation]

Everyone has their own culture of secrecy ... why release data and then get hammered for it in public ... there's a reflex of prudence ... I don't want to be at the back of the class.

For one participant, the culture of secrecy and fear of criticism are not the only factors that explain the reluctance of agencies to liberate organizational performance data.

[translation]

Beyond the culture of secrecy, a culture of measures for improvements through results-based management has yet to be developed.

Also cited was the difficulty of agencies to include real measures of impact in their processes of management and review of programs and services. [translation] "We put much more emphasis on having to justify any potential cost overruns than on explaining why we haven't met out targets."

2.6 SALE OF DATA

Agencies have a tendency to retain data that highlights their operations and performance, unless the government requires it to be released, as in the case of the Quebec government's electronic tendering system and its Dashboard on Health Projects in Information Resources. However, not all government data is held onto by agencies. On the contrary, one need only look at agencies' websites to note the plethora of information and documents of public interest. Some agencies, such as the SAAQ, have become experts in the dissemination of statistics, among other things, to increase public awareness of certain issues and encourage people to behave more responsibly. For example, "when it comes to road safety ... it is in everyone's interest that people become aware and take ownership of the issue."

However, the dissemination of data may hold commercial value and data may be sold by the agencies themselves, more often than not to external users.[15] Through the sale of data, agencies seek to cover, in whole or in part, the costs of acquiring, preparing and disseminating the data in a cost recovery context. For example, section 12 of the *Act respecting the Ministère des Ressources naturelles et de la Faune* [16] authorizes the department to "supply, in return for payment, geographical information." The development and dissemination of the most complete databank of addresses in Quebec are also part of this commercial approach.

---

[15] Data sold by agencies may sometimes be provided free of charge to other agencies, as in the case of ACRIGéo (Approche de coopération en réseau interministériel pour l'information géographique) where there is a free geographic information-sharing zone between 22 departments and agencies of the Government of Quebec. Some of the information has been commercialized outside government.

[16] Renamed the Ministry of Energy and Natural Resources following a change of government in 2014.

[translation]

Adresses Québec ... is sold at market cost … it was a request in open data. We understand that it can help support Quebec's economic development, but we would not be able to supply it if we were not self-funded.

Sales revenue can also be used to fund services seeking to prepare the data upstream (e.g.: standardization, cleanup, integration, anonymization and indexation) or to add value to it.

[translation]

Our bank contains 300 indexed fields ... We redact ... we mix and match various formats ... Until very recently, we were still receiving WordPerfect documents ... We have systems that are capable of extracting certain metadata ... Decisions of value are identified and then a whole editorial process is put in motion ... Summaries are done ... So, there is a huge amount of work that is done upstream upon receipt ... that information doesn't come free ... it comes at a significant cost ... there are 120 people working on it.

In the opinion of the agencies' representatives who sell data, if the government wishes to free up and make data that generates revenue available for free, it should plan on implementing measures to compensate for the loss of revenue absorbed by those agencies.

[translation]

For us the main issue ... is pricing ... To be sure, if there was suddenly no problem of self-funding, we would free up the data!

3. CONCLUSIONS AND POTENTIAL SOLUTIONS

In Quebec, government agencies already disseminate much information of public interest (e.g.: annual management report, populations and services statistics report and research report) on their websites and in various documents. However, much of this data is aggregated and contained in tables whose formats are difficult to reuse. Freeing up this data so as to render it reusable would be a step in the right direction.

However, for real benefits to accrue from transparency, agencies and government will need to go further in freeing up data that sheds light on their way of operating, the effectiveness of their programs, the efficiency of their services and many other strategic drivers of State performance. As our study shows, the liberation of data, particularly operational or transactional data that can be used for accountability purposes, runs into practices, systems, a culture and a legal framework that are more inclined to control information than disseminate it.

In this last section, we suggest levers the government and its agencies can use to encourage and make it easier to free up data useful to the development of applications and services and to promote greater transparency in government.

3.1 CENTRAL GOVERNANCE WITH CLEAR DIRECTIONS

For a number of participants, particularly open data respondents from the organizations we studied, the data liberation movement needs not only an initial political momentum to start, but also a government framework to direct and support it. The movement will take off in the public service when the government and central authorities provide [translation] "clearer *directives*" to agencies, we were told, accompanied by a vision and an implementation plan. [translation] "As long as it's not coming from above, I will not move," stated one participant. Without policy, direction, and an action plan from the government,[17] there is every reason to believe that agencies will continue to place their priorities elsewhere, given that they are already busy making sure they maintain current operations in a budgetary context that greatly limits their available resources. [translation] "As long as Treasury Board policy does not push harder ... if I'm not given any directives ... the more sensitive data, we will say 'no'." According to some participants, central authorities would be legitimized to designate programs whose data is to be liberated in the interest of transparency, particularly in the wake of the work done by the Commission and the Bureau de la révision permanente des programmes within the Treasury Board Secretariat.

Strong central governance in data liberation does not mean an absence of collaboration, on the contrary. Such collaboration between government agencies (e.g.: Treasury Board, departments and other provincial agencies, municipalities, Crown Corporations) is essential in order to produce and disseminate open data,[18] but it is also required between governments and societal actors (e.g.: citizens, NPOs, associative

---

[17] In much the same way as the Government of Canada and major member countries of the Open Government Partnership.
[18] At the same time, [translation] "it's a bit difficult for Treasury or for central agencies to go into detail in organizations to say: You need to liberate such and such dataset, because of all the constraints and consequences that would entail."

groups, developers, businesses) to develop government strategies and to (re)use data for development and transparency purposes. In that regard, it should be noted that "Canada has developed its [latest] national Action Plan in consultation with citizens, civil society organizations, and the private sector."[19] Canada has also issued a directive on open government that, in addition to supporting an open by default policy with regard to operational data, [translation] "provides clear and mandatory requirements that all departments must meet" and which are aimed at, in particular "any new plans for procuring, developing, or modernizing departmental information applications, systems, or solutions in support of the delivery of programs and services."[20] Such government directives must be accompanied by operational tools to guide agencies through the process of data liberation.

> *We were told: "You will need to free up data," but we were not told how, as if it was ultimately something simple ... there was an announcement made by a minister, but still no policy or framework.*

In order to guide agencies through the process of opening their data, the open government team of Quebec's Treasury Board developed and provided agencies with an operational framework for freeing up high-value open data.

An agency producing open data will also have an important role with respect to governance within its organization. First, a formal and stated support from the agency's senior management is essential for mobilizing the necessary resources and the organizational actors affected by data liberation, including those who are responsible for business lines, information technologies, access to information, communications, and many other stakeholders. Second, the agency must be equipped with mechanisms (e.g.: strategies, committees, standards, procedures) to effectively coordinate and organize the liberation of its data while complying with ethics laws and regulations, particularly with respect to the *protection* of sensitive personal information.

> [translation]
>
> *We need more structure and all relevant sections need to be involved ... Communications ... document management, legal ... Our directorate does not have the authority to open that data.*

Lastly, it is by integrating the data opening process early in the development cycle of information systems, rather than at the end of this cycle, at the moment the data is disseminated, that agencies will be able to aspire to the open by default principle, proposed by the G8 Open Data Charter and adopted by the member

---

[19] Government of Canada (2014), Open Government Action Plan 2014-2016.

[20] Government of Canada (2014), Directive on Open Government, in effect since October 9, 2014.

countries of the *Open Government Partnership,* which includes Canada, France, the United Kingdom and the United States.

3.2 POOLING EXPERTISE AND RESOURCES

There are many agencies that produce data that is of public interest. Some of these have large databanks on the programs and services they administer and on the citizens they serve. However, most of this administrative data, as with much government data, was not produced to be released and, even less, to be reused. Considerable work in preparing and formatting is therefore required to ensure that any reuse respects the confidentiality of sensitive information. In order to render Open Data intelligible, it will need to be accompanied by metadata that precisely define the content, format, and when it was updated.

[translation]

There needs to be coherence in what everyone produces ... If you don't enter your metadata, you might think there are several numbers for the same concept, the same variable ... the metadata is the intelligibility of the data. If it's not provided, the user will not have the right information.

By assigning a precise meaning to open data in a particular format, the metadata can contribute to the standardization of this data and, at the same time, facilitate its (re)use, whether to analyze it or to combine it with other data. To that end, it should be noted that much of the Government of Quebec's geospatial data has been standardized and formatted so as to make it easier to file and reuse it in Quebec's open data portal.

In contrast to geospatial data, the methodologies, standards and tools used to produce transactional or financial data may vary from one agency, and even information system or line of business, to another. For a number of participants, standardizing this information is necessary in a context of integration of services and government transparency. Some pointed out that this standardization needs to be done by means of a normative framework to structure metadata and other methodological elements in the management of information [translation] "no matter which source it comes from."[21]

Data liberation may also require the data to be rendered anonymous prior to being freed up, especially when it is a matter of transactions between the State and its citizens, by redacting any information that would enable identification of the individual's identity. Anonymizing data may mobilize significant resources, depending on the volume and nature of the personal information to be processed.

[translation]

We have a redacting team of about 15-20 people ... that's

---

[21] [translation] "The normative framework indicates how to set up your information system … At the end, you have the recipe … the objective being to arrive at unique information."

100,000 decisions per year … We also have robots that go over it … there's a part that's automated … but ... it requires verification by an actual person.

Lastly, the open data available on government portals and its metadata are above all raw material for developers or designers of applications. These are the primary reusers. As for ordinary citizens and other potential users (e.g.: journalists, researchers, NPOs, businesses and public agencies), open data on portals more often than not resembles a bunch of long series of unintelligible numbers. To reach these users, open data needs to be accompanied by tools to facilitate access and processing. Put another way, the democratisation of open data requires the development of visualization interfaces that are user friendly and (inasmuch as is possible) free in order to enable users to consult and search open data. Developed by actors from the private, public or parapublic sectors, a number of these tools can be found on government portals, including the new *Data Quebec* portal created in April 2016, under the 'Applications' heading.[22] Thanks to these tools, it is now possible to consult and search, for example, contracts awarded by all Quebec public agencies, the list of contractors that are banned from public contracts awarded by the Government of Quebec, occupancy rates in Quebec emergency rooms, food establishment convictions, the list of the most popular given names, and much more data. The number of such visualization tools tends to increase as more and more open datasets of public interest are made available.

Data liberation is therefore more than the mere publication of information on a portal. It presupposes an initial preparation of the data (e.g.: validation of compliance, anonymization, indexation and standardization) to facilitate and secure its reuse followed by a visualization of the data to ensure widespread use. In this regard, Quebec's public administration can count on agencies that have, over time, been able to develop expertise in standardizing and indexing open data (e.g.: metadata), as well as in anonymizing and integrating the data on a same site or consulting and searching that data. This is the case with the Institut de la Statistique for data on Quebec's population, the Régie de l'assurance-maladie and the Ministère de la Santé et des Services sociaux for health data, the Ministère de la Sécurité publique and the Ministère de l'Énergie et des Ressources naturelles for geomatics data, and the SOQUIJ for legal information. It is therefore in the government's interest to include these partners in the data liberation process.

---

[22] Site consulted in May 2016, https://www.donneesquebec.ca/fr/.

CONCLUSION

Open data has upended institutionalized practices in information management. It has in fact challenged the State's bureaucratic way of operating, characterized by a hierarchical control of information. Being used to protecting and controlling its data in predictable operational contexts, government is now invited to free up part of its data for (re)use, the concept of which is alien to it. In order to do so, it must adopt a network-based logic for action, the opposite of a hierarchy, which is characterized by seamless exchanges and a diversity of collaborations and contributions. This involves changes in the way government manages information and develops its systems, and in the way it deals with citizens, researchers, the media, developers and other societal actors. In freeing up its data, government is thus called upon to give up its role as a guardian of information and take on the role of promoting usable data, first, to facilitate the development of applications and low-cost, value-added services, and second, to account for its performance and probity.

This study shows that various organizational and technological conditions tend to hinder the liberation of open data in the Government of Quebec. However, such impediments should not be viewed as insurmountable obstacles that are irreconcilable with the principles of government openness and transparency. Paradoxically, the use of certain hierarchical and bureaucratic mechanisms can be an effective way of reducing such obstacles. In fact, representatives of agencies we met with expect government authorities to provide them with clear directives and practical tools before moving forward with data liberation, particularly in terms of transparency. There is every reason to believe that agencies will not undertake such a shift by themselves in a budgetary context that leads them to focus their efforts on maintaining current operations. As a bureaucratic mechanism, the standardization of data through the use of consistent methodologies and metadata has also proven to be effective in ensuring the uniqueness of the information and, at the same time, facilitating the (re)use of the data.

If the liberation of government data could benefit from being centrally coordinated and standardized, *voluntary* cooperation of agencies cannot be decreed by a central authority. Such collaboration is more easily obtained when agencies draw real advantages from data liberation through conclusive experiences with (re)use. Although such experiences remain few and little-documented, the government can count on a certain amount of goodwill from agencies. Indeed, participants readily acknowledge that open data can be advantageous to agencies that wish to 1) have applications and services developed at a low cost, 2) compare themselves to other agencies or 3) use government information, from here and elsewhere, to support them in their strategic decisions without having to make access to information requests (see section 1.3).

In terms of geopositioning, a number of agencies are already benefitting from sharing their data thanks to the creation of a common visualization platform, known as the Infrastructure Géomatique Ouverte (IGO).[23]

---

[23] [translation] "IGO was developed through a community of geomatics specialists from several ministries and agencies in Quebec's public administration. It allows for the use of a multitude of geographical data (cultural heritage, agricultural lands,

The platform enables an agency to geographically locate its data as well as data from other agencies through an open mapping interface. The agency can then use the data from a network of contributors [translation] "at a fraction of the cost of acquisition." For an agency that wishes to familiarize itself with the mechanics of data liberation or for other reasons, it is possible for it, through the IGO, to first open its data to government agencies before making it public. [translation] "If a department decides to transfer a dataset that is in an intranet zone to the public, it just has to press a button and there is nothing to change."

Lastly, experimentation with promising or conclusive initiatives, conducted by government and non-government actors, has rendered the phenomenon of data liberation concrete. The promotion of these initiatives and their marketing, combined with political and administrative leadership that provides clear directions, can pave the way for the public service to open data with a view to improving services and government transparency.

---

schools, fire halls, derelict villages, etc.) thanks to a mapping interface accessible through a web browser." (IGO website, consulted in May 2016, http://igouverte.org/)