

Driving economic and social growth

Designing an open data strategy for Public Sector organizations

Content

The rise of open data	1
Open data in Canadian public sector organizations	2
Creating an open data strategy	3
The economic and social value of data	4
Additional steps to be taken to maximize the value of datasets	6
Funding models	7
Access to data	8
Getting started	9
How Deloitte can help	11

The rise of open data

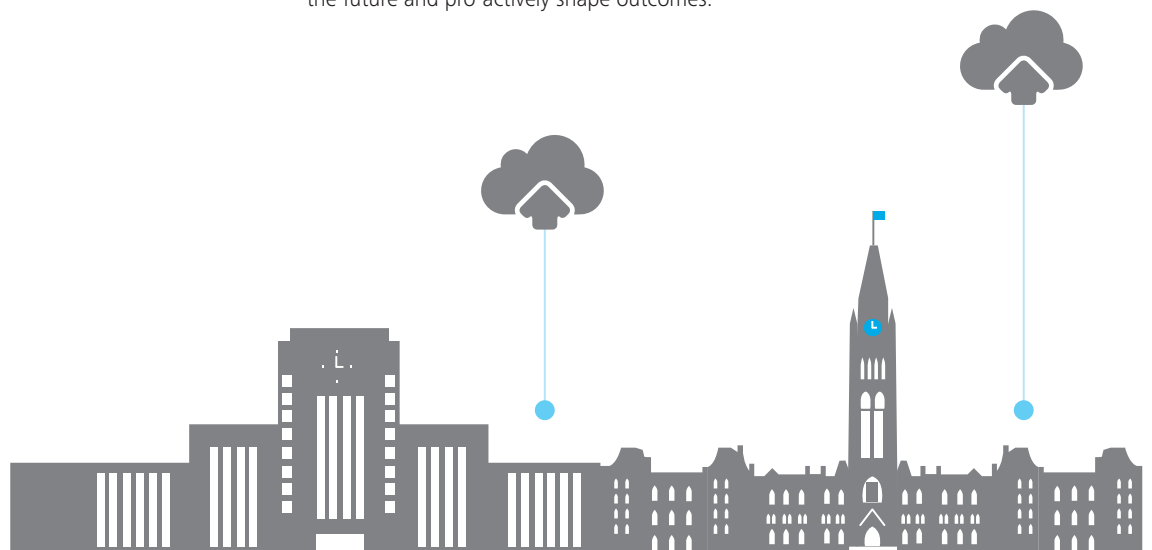
It has been said that 'data is the new oil'.¹ However, unlike oil, data will not eventually run out and has a history of being made available at no cost to end users. Perhaps a better analogy is that "data is the new carbon atom" – something that is truly abundant and can be refined further to create even more value.² Like the carbon atom, the real value of data comes not just from its generation, but from its manipulation and analysis. By doing so, policymakers and commercial organizations can uncover insights that improve outcomes and yield financial benefits. Recognizing this, across the economy, both the public sector and businesses are increasingly investing in processes and technologies to maximize the value of the data they are collecting and generating.

Open data³ is increasingly important. Whereas initial moves to release more open data revolved around improving government transparency to promote accountability and better public service delivery, the ongoing provision of open data can go even further, exploiting opportunities across at least six areas. Ongoing benefits of releasing more and higher quality open data include:

- **Increasing accountability:** using data to hold politicians and public bodies to account;
- **Strengthening public services:** using data to compare and contrast services and improve decision-making, which in turn can also drive service improvements;
- **Investing in what works:** using data to benchmark activities and make efficiency improvements and better strategic decisions;
- **Engaging citizens and service providers:** data can help increase citizen engagement and allow for better alignment of incentives of service providers;
- **Stimulating social growth:** using data to improve the quality of public discourse; and
- **Driving innovation and facilitating economic growth:** using data to stimulate innovation and ultimately job and wealth creation.

The most forward-thinking businesses have already been able to exploit insights from multiple datasets to understand better the reasons behind past performance and to predict trends in consumer behaviour – though even limited data analysis can improve business performance and profitability. This transition from hindsight to foresight is allowing businesses to plan for the future and pro-actively shape outcomes.

Even limited data analysis can improve business performance and profitability



Open data in Canadian public sector organizations

The public sector is also at the forefront of change. Across Canada, federal, provincial and municipal governments and broader public sector (BPS) organizations are beginning to develop and publish open data strategies that set out the steps to be taken to embed best practices around the collation, use and dissemination of public sector data.

With the assistance of organizations like the Canadian Open Data Institute, early pioneers are already seeing the benefits. Examples include:

- **The Government of Canada's *Open Data Portal*** (data.gc.ca) which launched in June of 2013. It has already grown to include over 191,000 datasets across 35 government departments and agencies.
- **Revenue Québec**, Quebec's tax collection department, has demonstrated the value that unlocking the power of data can have for the public sector's bottom line. *The Wealth Indicators Project* used a cross-section of government and external data to estimate the expenses of individual citizens to maintain their lifestyle. Comparing that information to tax filings provided an indication of potential tax avoidance that would otherwise go undetected. From 2003-2011 the government was able to collect an additional \$189 million in revenues as a result of the project, for a total **fiscal benefit to government of \$157 million** after expenses.
- **The City of Vancouver** has developed *VanMap*, a Web-based application that lets residents and others view data about Vancouver in map form. *VanMap* combines data from a wide range of municipal datasets and from other data sources including the federal government. Using *VanMap*, one can view property lines, zoning information, sewer mains, water mains, information about addresses, crime data and more at no cost.
- **The City of Toronto** has released a wide range of datasets, and enterprising mobile app developers have already created over 20 applications that use City of Toronto data. One of the open datasets is comprised of real-time information about the location of public transit (TTC) vehicles. Based on this data, several developers have created popular applications to help Torontonians plan their daily commutes. The *localMotion* app pulls together open data on multiple transit options including

local public transit, regional public transit, bike sharing programs and car sharing services.

- **The Government of British Columbia** has created a new department, *DataBC*, with a mandate to make a wide range of government data available at no cost. To date, *DataBC* has made 2,500 data sets available such as birth rates, old-growth forestry management areas, carbon emissions statistics and information on schools. This information is searchable and available for anyone to use. The goal is to help citizens and businesses make informed decisions, conduct research, analyze statistics and develop applications. These efforts are already paying off:

– *The School Zone* app helps parents and students identify the most dangerous intersections to find the safest route to school. It was developed during a weekend "hackathon" using datasets of school locations from the Ministry of Education and crash statistics from the Insurance Corporation of British Columbia.

– *HectaresBC* merges Government of BC data with environmental non-profit data, and allow any citizen regardless of their expertise to search, study, and understand the rich eco-diversity and land management challenges confronting the province.

These open data initiatives are beginning to attract positive momentum amongst Canadians; for example, Statistics Canada reported a doubling of data access since the removal of their licensing restrictions in February of 2012, in addition to positive feedback from diverse user groups, including academics, economists, journalists and financial analysts.

However, all levels of Canadian government and the public have yet to unlock the full capabilities of these resources, with Canada ranking only 10th on the international Open Data Index, which rates countries on the availability of data in 10 key areas⁴. In the global competition for business investment and top talent, Canadian government organizations can promote growth and investment by tapping into (and enabling others to tap into) their data assets.



Creating an open data strategy



Whether in Canada or globally, and at any level of government there are a number of key issues that public sector organizations need to address when developing and implementing open data strategies. Based on our experience in advising and working with public sector and private sector organizations to develop and implement open data strategies, Deloitte has identified a number of crunchy questions for public sector organizations that will inform and help to shape open data strategies.

These include:

1. **What is the expected economic and social value of the dataset in question? What is the underlying value chain? In other words, how is it expected to be used or re-used?**
2. **What additional steps need to be taken to ensure the data can generate the greatest value for the greatest number of citizens and businesses?**
3. **What is the most appropriate business model to fund the collection and dissemination of the dataset? Can the dataset be made available as open data?**
4. **How should users access the dataset? Are there privacy, security or copyright issues to be overcome?**

Failure to adequately understand and address these issues from either a supply or a demand-side perspective could result in value being lost and the full potential of open data not being maximized.

The economic and social value of data

The value of any given public sector dataset can be disaggregated into four distinct components:

- **The direct value of public sector data accruing to its producers and suppliers:** these are the benefits accruing to producers and suppliers (i.e. those public sector organizations that generate, collect and disseminate data – also known as public sector data holders) through the sale of public sector data or related value-added services;
- **The indirect value of public sector data arising from its production and supply:** the benefits accruing up the supply chain to those organizations interacting with and supplying public sector data holders, and the benefits accruing to those organizations where employees of public sector data holders and supply chain organizations spend their wages;

- **The direct use value of public sector data to consumers of public sector data:** the benefits accruing to businesses, civil society, individuals and the public sector from directly using and re-using public sector information for a variety of purposes; and
- **The wider societal value arising from the use and re-use of public sector data:** the benefits to society of public sector information being exploited, which are not readily captured elsewhere.

The first three types of value can be grouped as narrow economic value and can be measured using standard economic methodologies such as willingness to pay surveys and input-output analysis to derive a monetary estimate for value of each dataset and the associated expected number of jobs it could support.

Driving growth, ingenuity and innovation

There are a number of ways data can be exploited to generate value for public sector and private sector organizations. Especially with large datasets, data analytics can be employed to discover meaningful patterns, test hypotheses and predict outcomes in an increasingly sophisticated manner. On a practical basis, analytical techniques can include a combination of:

- **Data Mining:** extracting insights from large datasets to detect anomalies and data clusters, as well as relationships between variables and summary statistics and profiling (data dictionaries);
- **Network Analysis:** charting nodal relationships between multiple variables over time and by geographical area;
- **Hypothesis Testing:** using a range of statistical techniques (such as regressions analyzes, non-parametric methods, decision-tree modelling, etc.) to validate theories and identify quantitative relationships between variables (i.e. moving from correlation towards causation);
- **Prediction:** using a variety of techniques to use historical data and contemporary understanding to predict future outcomes or ranges of outcomes; and
- **Optimization:** taking the insights of prediction and hypothesis testing and embedding them in business decisions (often in an automated manner) to help clients meet their objectives.

Through data analytics and other methods, public sector data can facilitate the development of new products and services built directly on public sector data; reduce transaction costs; and generate efficiency gains in the public sector itself.

The final type of value is harder to measure as it captures wider benefits to society arising from the use of public sector information – these are typically not measured in monetary terms. The literature discusses a number of ways in which public sector data can have broader impacts:

- **Increasing democratic participation:** giving citizens and businesses access to public sector data allows them to perform their own analyses of salient issues, make more informed choices about public service providers and interact with policymakers to challenge their assumptions and improve the policy making process;
- **Promoting greater accountability:** for example through the scrutiny of costs of public service provision and benchmarking comparable services;
- **Greater social cohesion:** for example, by providing more information on the provision and distribution of services, public sector information can be used to dispel myths on who receives certain public services;
- **Generating environmental benefits:** such as reducing congestion and pollution through the release of better traffic and transport data which helps drivers to better plan journeys; and
- **Identifying previously unknown links between different policy areas:** through data mash- ups it may be possible to develop system-wide solutions that holistically seek to address the root of policy challenges.

The social value of a particular dataset can be estimated using bespoke techniques and proxy measures, such as those used to estimate the economic impact of public sector information for the Shakespeare Review in the United Kingdom.⁵

Of course, one of the biggest challenges in estimating the value of any given data is being able to predict how it might be used in the future – the use of an open dataset is not always obvious at the outset. It may be the case that datasets with hitherto little perceived value may in the

future be the source of significant additional benefits as users discover new ways to use them⁶. Equally, a previously valuable dataset may no longer generate the same amount of benefit if it is superseded by another product using different data. In such instances, it is important that the estimates are constantly monitored and reviewed to update in light of new information. There are steps one might take during the initial estimation to identify how value might change, for example by considering how the value of a particular dataset may differ if it was linked with other datasets.

There will be a different value chain for each dataset and it is important that public sector organizations are able to identify the key components of the value chain – by doing so they can build robust cost-benefit cases that will inform how much resource to dedicate to collecting/refining/ disseminating the dataset and what business model to use to fund it.

Questions for public sector organizations

- Can a clear value chain be identified for each dataset setting out who is generating the dataset, who is disseminating it, who the likely users are and how it will be used?
- Can a monetary estimate of the economic and social value of the dataset be undertaken that will capture both current and future value? This can inform a cost-benefit analysis and an assessment of the most appropriate business model.
- Can public sector bodies ‘brainstorm’ what future uses of data will be with ideas from the wider community to help identify future value?
- How much data is being generated by the public sector data holder as ‘exhaust data’ (also known as administrative data which is data generated as part of its daily activities at little or nominal cost)?
- How much data is being collected as part of its public service duties?
- Which datasets are not being disseminated and why?

Additional steps to be taken to maximize the value of datasets

While it may be the case for many datasets that users can immediately begin to extract value, for other datasets a degree of refinement by the public sector data holder or another body may be required before the data can achieve its full potential. Equally, there may be wider supporting policy initiatives that the public sector can undertake to ensure that the widest possible audience has the necessary skills sets to exploit value from its datasets.

Some of the additional steps that the public sector data holder can take include:

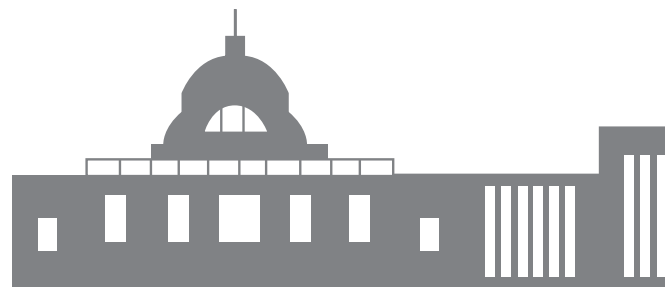
- Ensuring datasets are fully compliant with the Open Government Licence (or equivalent) on conditions of usage and do not contain any further restrictions on use other than, at most, a requirement to attribute the data to the original data holder;
- Ensuring the dataset complies with open data standards around machine readability and can be used with the widest number of software packages (including open source options);
- Having metadata files explaining the datasets and having named contacts and details of staff in public sector data holders that can answer queries;
- Being clear which datasets will be collected and made available on an ongoing basis, and which datasets are purely experimental and one-off;
- Where possible move datasets towards the five star level of open data, as set out by Sir Tim Berners-Lee,⁷ which includes linking datasets to one another to provide greater context;
- Making the data available in both 'raw' format for expert users and in dashboard format for casual users, or to promote transparency; and
- Developing skills strategies to ensure users have the ability to extract the most value from datasets.

Of course there is a balance to be reached here. This is an incremental process and the public sector needs to be aware that in many respects this remains a nascent market and so any large scale initiatives will have a long-lasting imprint on how the market evolves. For example, while providing highly refined datasets in easy-to-access formats, such as dashboards, may generate value today, they can risk crowding out private sector providers who could have built such services themselves and innovated around them. Similarly, there will be little public value in moving datasets to the five-star level if there is little demand or expected value.

When considering what supporting steps to take, reference should be made back to the expected value of each dataset and who the audience is likely to be.

Questions for public sector organizations

- What are the supporting activities, public sector data holders can take to facilitate end users extracting value from datasets?
- Is there an economic rationale for providing public sector support for public open datasets? In some cases, activities to maximize value today may not represent long-term value for money.



Funding models

One of the most controversial policy questions around public data is over the business model to be used to fund a dataset's collection and dissemination. The arguments over whether datasets should be made available at no cost to end users (i.e. as open data) or for a fee have been explored in detail elsewhere.⁸ Broadly speaking the decision whether to make datasets available at no cost to the end user or for a fee will depend on the answer to the following questions:

- What is the cost of generating/collecting/refining/ disseminating the data and what is the expected economic and social value expected from making it more widely available either as open data or pay-for data?
- Would the loss of any current revenue from pay-for data, and the costs of providing additional data sets for free, be recovered through other means such as increased tax revenues elsewhere or reduced costs?
- Is the dataset in question of national importance and public good? Is the funding for this dataset protected or could it conceivably be reduced in the future?
- Are additional supporting activities required?
- Is a charge for the dataset acting as a barrier to entry for small businesses?
- What is the ability to pay of users? Is there scope for discounts for certain types of users?

These are not straightforward questions to answer and it is likely that there will be different responses for different datasets. What is of critical importance is that there is transparency behind the choices and that these issues are regularly revisited to see if circumstances have changed. In some cases, there may be an element of quid pro quo in that users of data need to share, subject to protecting their intellectual property, how they are generating value from certain datasets or intend to do so.

Questions for public sector organizations

- What is the economic cost of generating and disseminating data? Is this outweighed by the wider economic and social benefits arising from its use?
- What funding models are available for open data? Importantly, there is not always a binary choice – organizations can choose to release 'freemium' products or move from a position of pay for data towards open data over time. For example, while initially data may carry a fee due to high costs of refinement, over time these processes could be automated or users become able to do refinement themselves, meaning the underlying data can be released as open data
- If there is not a clear monetary estimate of the benefits of open data, can the case for its release for no fee be clearly articulated? Can a fee be levied on the data at some later point if justified?



Access to data

It should be remembered that there will be limits to which datasets can be made wholly available as open data irrespective of cost. Certain datasets contain sensitive personal records of individuals and are typically restricted to authorized users. However, such datasets, for example in health and education, often yield the greatest societal value and public organizations should consider how they can safely increase access to them. Increased access can be achieved through a number of avenues including:

- Creating safe environments (physical or virtual) for authorized users to access the datasets but not be able to take away the underlying data;
- Vetting potential users and having signed declarations for users to access the datasets; and
- Releasing edited or anonymized and aggregated datasets that do not contain personal data or record-level information.

If the answers to the first two bullet points are that safe environments and vetting are necessary, then the data is unlikely to be made available as open data.

If the data can be released in anonymized or disaggregated form, one of the key issues for public sector data holders to consider is whether such an anonymized or aggregated dataset can still yield comparable value to a non-aggregated dataset. Even if it does not, there may still be value in making this version of the dataset more widely available.

Questions for public sector organizations

- Is the dataset in question suitable for release as open data – are there any issues around privacy and security?
- If so, can access still be provided to researchers and commercial firms in order to yield benefits?



Getting started

Guidance for developing an open data strategy for your organization

Given the range of issues to consider, developing an open data strategy can appear daunting to government organizations of any size or mandate. A structured approach can help make the process more manageable and indicate the level of resources required for strategy development.

Step 1: Identify data sets

Do not try to make all of the organization's data sets available at the outset. Focus on those data sets that are of high value to citizens and stakeholders and that can be shared simply and inexpensively. Create a community of users and tap into that community to help identify what other data might be useful, and why. Work collaboratively with partner organizations, departments or ministries to develop a common framework for identifying these "low-hanging fruit."

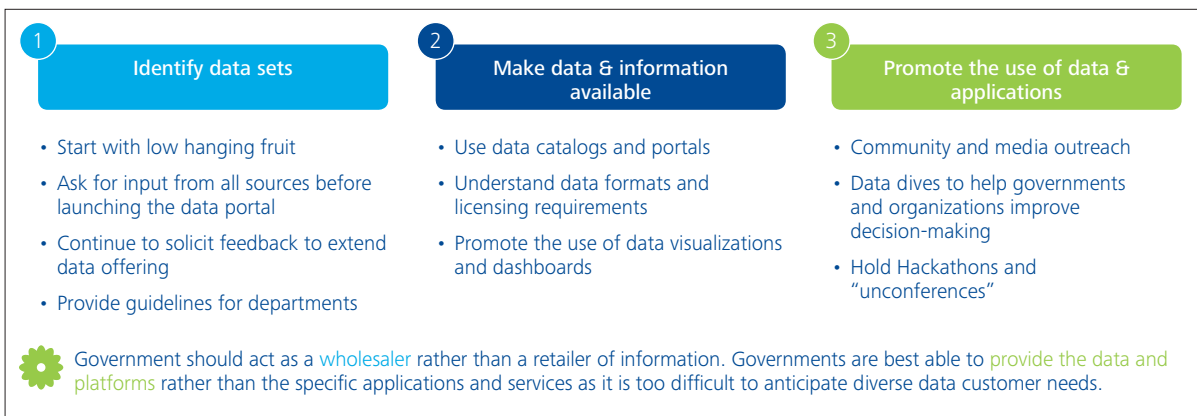
Step 2: Make data and information available

Data will not be transformed into public value unless it is accessible to users. Use data catalogs and a data portal to make it easy for users to find, download, manipulate and visualize the data. Remember that government organizations should focus upstream, on the data and platforms, not the specific applications. Leverage the community of developers to determine and meet citizen/consumer need.

Step 3: Promote the use of the data and applications

Generate value by convening the community of data users and application developers. "Hackathons" and other events can generate interest and awareness, catalyze the ecosystem of stakeholders and deliver meaningful results for citizens.

Figure 1: Creating an open data strategy for government organizations



Source: Deloitte, 2014

How Deloitte can help

This paper provides an overview of the importance of open data, highlights a number of examples of open data initiatives being pursued across the country and internationally, and discusses a number of the issues that need to be considered when starting to formulate an open data strategy.

As discussed above the concept of open data is gaining increasing traction around the globe and is not going away. If they have not already done so, all levels of government will quickly need to engage on this topic and formulate their own approach to avoid falling further behind and ultimately having an approach forced on them through public pressure.

The Canadian authors would like to recognize the significant content input from Haris Irshad, Harvey Lewis and Andy Tong from Deloitte UK.

Deloitte is able to draw on our extensive experience in this area both within Canada and internationally to help you formulate your open data strategy, work through the issues involved and develop an implementation approach that ensures success across all three phases of a comprehensive open data initiative: Pre-Data Release, Data Release and Post-Data Release.

For further information on this topic and to discuss how we may be able to help your organization please contact a member of our *Public Sector Transformation Consulting* team.



Andy Potter
andpotter@deloitte.ca
416-643-8938



David Stewart
davstewart@deloitte.ca
416-775-7484



Alex Case
acase@deloitte.ca
416-601-6408



Charles Perron
cperron@deloitte.ca
613-751-5243

For more information visit www.deloitte.ca/analytics OR contact us for a demo of open data solutions at the Deloitte Greenhouse deloitteanalytics@deloitte.ca

www.deloitte.ca

Deloitte, one of Canada's leading professional services firms, provides audit, tax, consulting, and financial advisory services. Deloitte LLP, an Ontario limited liability partnership, is the Canadian member firm of Deloitte Touche Tohmatsu Limited. Deloitte operates in Quebec as Deloitte s.e.n.c.r.l., a Quebec limited liability partnership.

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee, and its network of member firms, each of which is a legally separate and independent entity. Please see www.deloitte.com/about for a detailed description of the legal structure of Deloitte Touche Tohmatsu Limited and its member firms.

© Deloitte LLP and affiliated entities.
Designed and produced by the Deloitte Design Studio, Canada. 14-2281