HOUSE OF COMMONS
CHAMBRE DES COMMUNES
CANADA

Standing Committee on Government Operations and Estimates

EVIDENCE

Thursday, May 15, 2014

—

Chair

Mr. Pierre-Luc Dusseault

Standing Committee on Government Operations and Estimates

Thursday, May 15, 2014

● (0845)

[*Translation*]

The Chair (Mr. Pierre-Luc Dusseault (Sherbrooke, NDP)): Order, please.

Welcome, everyone, to the 26th meeting of the Standing Committee on Government Operations and Estimates. We are continuing our study on the government's open data practices.

Today, we are fortunate to be welcoming representatives from various Government of Canada departments and institutions. They will give brief presentations to explain how they manage open data practices and the needs of users. I won't introduce them all now, but rather will introduce them one by one as I give them the floor.

We will begin with the representatives from Statistics Canada, Bill Joyce, director of the Operations Branch, who is accompanied by Yves Éland.

Without further ado, I give them the floor for five to 10 minutes, depending on what they have to say this morning.

[*English*]

Mr. Bill Joyce (Director, Operations Branch, Statistics Canada): Mr. Chair, committee members, I thank you for the opportunity to address the committee today.

This short presentation will cover three areas. First, I will briefly explain how Statistics Canada is adopting the principles of open data in our dissemination practices. Second, I will explain how our agency is contributing to the Government of Canada open data portal. Finally, I will explain our role as a service provider for data. gc.ca.

Clearly the concept of open data is a natural fit with our agency's mandate. Providing our data free of cost, free of barriers to redistribution, and the machine-readable formats allow us to make our output more accessible to data users. Over the years we have steadily moved to increase the amount of free statistical information found on the Statistics Canada website, which is our main dissemination vehicle. In February of 2012 our organization took an important step by making all standard data on the web available free of charge. We also adopted an open licence framework and eliminated all royalty fees, even for custom data tabulations.

Since then visits to the Statistics Canada website have increased by over 20% and web traffic to the CANSIM application, which is our main output database of socio-economic data, has doubled for that application.

I will now move on to talk about our contribution to the Government of Canada open data portal. It is our approach that all standard aggregate downloadable data that are found on the Statistics Canada website should also be discoverable through data.gc.ca. To date, we have registered 5,400 data sets to the portal, which represents approximately 72% of the general non-geospatial data sets.

The data sources include census population, the national household survey, census of agriculture, output from our CANSIM database, our summary tables, industry and occupation classification files, and most of our geographic reference files. Monthly international trade data will become available to the open data portal in the next couple of months, and more classification files and geographic files will also be loaded over the course of the next year.

We do have evidence that data users are accessing our data sets via the open data portal. From the CANSIM database alone, we know that about 6,000 download requests were generated from data.gc.ca.

I know you've heard about the recent Canadian open data experience, hackathon or codefest, in which students, entrepreneurs and developers focused on building applications with data found on the open data portal. During this event our data sets were accessed approximately 1,500 times, which was more than any other federal department. Of the 15 finalists in this event, seven used data from Statistics Canada.

These are some examples that demonstrate to us that our data sets are, indeed, reaching a new audience through data.gc.ca.

The last slide explains the role of Statistics Canada as a service provider for the open data portal. In addition to our role as a main data contributor, you are aware that the second generation portal was launched last June, but you may not be aware that Statistics Canada plays a role behind the scenes for the portal's system development and technical support operations. This service delivery is governed by a memorandum of understanding with Treasury Board Secretariat, which is our client in this endeavour.

Mr. Chair, this concludes my remarks. I'd be happy to take any questions committee members may have when you get to that point.

Merci.

● (0850)

[*Translation*]

The Chair: Thank you for your presentation.

I now turn to the representatives from the Department of Natural Resources, Mr. Shukle and Mr. Ferland, who will share the time allocated to them.

Mr. Pierre Ferland (Chief Information Officer, Chief Information Office and Security Branch, Department of Natural Resources): Thank you, Mr. Chair, and good morning to the committee members and my colleagues from the department.

My name is Pierre Ferland, and I am the chief information officer at NRCan. I am accompanied by Prashant Shukle, director general of the Earth Sciences Sector.

This is my first appearance before you, and it is a pleasure to have the opportunity to speak with you about our experience with open data and to answer your questions.

[*English*]

If you know the history of NRCan going to back to the origins of it in 1842, the creation of the geological survey commission has always been about developing information for purposes of dissemination to Canadians in businesses, so, for us, the concept of open data is a natural extension to that. It's a new channel, essentially. It started with maps and survey reports and it expanded to more information, so we've always been strong supporters of open data.

We recently participated in the CODE experiment, the Canadian Open Data Experience exercise, and one of our data sets, specifically, in this case, the vehicle fuel consumption, was used in collaboration with information from transport to create one of the top 20 apps. It was called CAN Fuel.

We publish open data on the topics of forestry, mining and extraction, and energy efficiency, but of course—as referred to by my colleague from Stats—one of our core products is in the geomatics area and maps, and I will let my colleague Prashant tell you more about this and our experience there.

[*Translation*]

Mr. Prashant Shukle (Director General, Canadian Centre for Remote Sensing – Geomatics Canada, Earth Sciences Sector, Department of Natural Resources): Thank you.

Mr. Chair and committee members…

[*English*]

first, let me explain that geo-data is the basic geographic or geoscience data that describes Canada's land mass. Some notable examples of this kind of information include geological information about where mines are and where you can find particular mineral deposits. There are topographic maps. For those of you who have cottages, you have probably used these topographic maps, which include data about things like water, lakes, roads, elevation, and all kinds of important data points that are becoming increasingly important in an open data environment.

The key distinguishing aspect of these data is that they are all defined by a location or position on the earth. Additionally, they are often relevant in multiple applications, ranging from property rights, to government policy decisions, to regulatory decisions, to environmental assessments, to estimating resource potential, and even right down to in-car GPS navigation, making sure that the pizza guy gets to the house within 20 minutes.

In the early days, the most useful form possible generally meant recording this kind of data on paper maps. Over time, we progressed to managing our geo-data holdings in NRCan as digital files on computers, although the final product was still paper maps.

I want to give recognition to Canadian leadership in this regard. Roger Tomlinson actually invented geographic information systems back in NRCan in 1964, so we've had a leadership role internationally and helped to spawn and create a multi-billion dollar industry worldwide.

Today, we make raw digital data, also known as machine-readable data, accessible over the Internet in forms that can be manipulated, combined, and transformed according to need. This is really at the hub of the concept of big data and big data analytics. Many departments and agencies make significant investments to collect and manage their geospatial information, but for various reasons, barriers exist that have prevented the timely sharing and integration of this information across the federal family and with partners. That's why we welcome the open data initiative.

Working together across government, the federal committee on geomatics and earth observations—the FCGEO, as it's called—is an ADM-level committee consisting of 21 departments and agencies working in very close collaboration with Treasury Board and the chief information officer, who's been directing overall government efforts with respect to open data under the guidance of Minister Clement. What we've been doing is working to break down barriers and capitalize on the full potential of the government's geospatial data and holdings.

● (0855)

[*Translation*]

The Federal Committee on Geomatics and Earth Observations, the FCGEO, consists of producers and/or consumers of geospatial data that are voluntarily collaborating in this broad federal effort. They have adopted an inclusive, open and transparent approach with each other in the interest of not just the federal government, but for the benefit of all Canadians. Recognition of the power and potential of new technologies and media, and of the importance of authoritative, open data to Canada's knowledge economy and global competitiveness was also key to moving this initiative forward.

Together, we are currently working towards a federal geospatial platform that will make federal government geo-data available through the government's open data portal, as well as support the sharing and reuse of data within the government.

[*English*]

With more than 10 years of experience in open geo-data, and one of the first public organizations in the world in the geospatial space to go open, NRCan's earth sciences sector has learned many lessons, although please remember our starting point that Pierre talked about. It goes back to 1842. It's been really hard work.

The work of surveyors and mappers really took a very structured and disciplined approach to how it is we collected the data, how it is we managed the data, and how it is we made it accessible. So there was a very highly structured approach that made it usable and actually facilitated the value-added dimension to the digital data we now produce.

NRCan has always intended that the geo-data it collects and manages be used by governments, industries, and citizens. Let me give you a practical example. The surveyors that went out in the old Department of the Interior, which preceded NRCan, actually helped us define our boundaries. It helped us shape our country, so it was absolutely essential that we communicated that to Canadians and to the world generally. It helped define our borders.

Coming back to the open data portal, NRCan was the principal contributor to the Government of Canada's open data portal pilot project when it was launched in March 2011 as part of the open government strategy. More than 90% of the available open data at that time was geospatial, originating from NRCan.

What has producing open data meant for NRCan? First, we've actually realized some savings because we don't need to have physical storage space or a vast distribution network to disseminate our physical products. We make our products, our data, available to such key players as Google and Google is able to disseminate our data in Google maps. It uses our data.

I understand you had Colin McKay before you a few weeks back. Colin would say that the partnership of Google and NRCan is one of the best partnerships they have between a federal government and Google.

However, there are new costs for maintaining servers, bandwidths, licences, and for uploading data files.

[*Translation*]

We have learned that accessible, free data are very much in demand. For example, geo-data downloads were less than one million a year in 2007 from our GeoGratis website and are now in the order of five million downloads last year, plus another two and a half million from our federal, provincial, territorial GeoBase website.

● (0900)

[*English*]

Just to recap, it was less than a million before we went open and between five and seven million after we went open, not of access points or hits, but downloads of our geospatial data. They're impressive numbers, but they're not downloads of interesting pictures or video clips. Most of these downloads are very large, complex data sets that are accompanied by detailed metadata. This means that they're most likely downloaded purposefully by someone who has the tools and the technology to manipulate the data and who sees potential benefit from reuse and packaging.

These data sets are complex in nature and as a result the file formats used are not always simple and not always open. We use open standards as much as possible, such as those from the International Organization for Standardization, more commonly known as ISO, or the international Open Geospatial Consortium, more commonly known as OGC, but we also make use of industry standards. Many historical products are only available as scans because that's all we have. Technically, it's not open data, but we make this information publicly available nonetheless.

While the download statistics indicate the geo-data are considered useful, the economic or social impact of geo-data reuse can be difficult to quantify. I also understand that you had a witness from the McKinsey group who also appeared and estimated the value, or at least in the American context identified it, if I remember correctly, as $3 trillion, if I'm not mistaken. Because it is open data, you don't really track the reuse potential all the way to its logical extension, so it's really difficult to quantify that number, but we're trying.

[*Translation*]

As a result, quantitatively understanding the clients and their level of satisfaction is difficult. Nonetheless, we have received much positive feedback from the clients about what we have available, but they usually want more. Interestingly, we have also seen an increased demand for older, historic products.

Conceptually, we accept that if the original data acquisition was judged to provide value for money, any additional reuse can only compound the benefits. To better understand the impact of open data, we are currently evaluating the impact of open geo-data in the marketplace.

[*English*]

Another area where we're learning lessons is in providing a simple way for users to easily find and access what they want. As more and more data sets become available through single portals, it becomes more difficult for the user to find that needle in the haystack.

In addition to the economic benefits of open data, our statistics show a lot of reuse of the data within the federal family. The ongoing broad-based engagement efforts of the Federal Committee on Geomatics and Earth Observations has been worth it. Current standards and approaches and those under development are the key to enabling accessibility and interoperability of the data and will enable future breakthroughs.

In closing, I want to reiterate that, from NRCan's perspective, our deliberate and intentional move towards open data was not simple, nor was it accomplished in the last few years. In fact, we've been working on it through most of our history, long before the Internet community introduced the phrase to describe the concept. Yet the journey in the Internet age has been definitely worthwhile, and we're beginning to see substantial benefits and new opportunities arising from our efforts.

Thank you for the opportunity to speak with you, and we'd be happy to take any questions.

[*Translation*]

The Chair: Thank you.

I now give the floor to Mr. Kiziltan, who is representing the Department of Citizenship and Immigration.

Thank you for being here today. The floor is yours.

Mr. Ümit Kiziltan (Director General, Research and Evaluation, Department of Citizenship and Immigration Canada): Good morning, Mr. Chair, and thank you for giving me the opportunity to present Citizenship and Immigration Canada's (CIC) participation in the open data portal.

[*English*]

As many of you might know already, CIC's data differ somewhat from much of the other data available on the Government of Canada's open data portal, due to the nature of our work. Using our data, we produce statistics, be they on permanent and temporary residents coming to Canada, grants of citizenship to new Canadians, or the processing of those applications. The personal nature of this information makes our data sets popular but also limits the amount and type of data we may make available. We take seriously our responsibilities to protect personal information and ensure the appropriate treatment of information. For example, the department aggregates our data to protect personal information, which I will speak about in a few minutes.

Based on the requests for CIC data that we received before we even began to post them on the open data site, and from current requests, we gather that parties interested in our data typically include prospective permanent or temporary residents, immigration consultants, lawyers, researchers, interest groups, corporations, members of the media, other federal departments, and provincial and municipal governments.

Treasury Board statistics show that while more than half of the clients downloading our data sets reside in Canada, a large number also are from abroad, with many of them in India, the United States, Pakistan, China, the United Kingdom, and Brazil. Given that our most popular data sets relate to permanent resident applications processed abroad, processing times, and permanent resident overseas inventories, we expect that many of our overseas clients are persons who have applied to come to Canada as immigrants or who are considering doing so.

● (0905)

[*Translation*]

The CIC data sets that are part of the open data initiative were originally released by the department on October 1, 2009, before the initiative was launched. These files, all part of CIC's *Quarterly Administrative Data Release* on CIC's own website, were disseminated at that time through an interactive interface that was published on a DVD and distributed via courier to anyone who requested one.

[*English*]

However, in light of and thanks to Canada's action plan initiative on open government to help the public find, download, and use Government of Canada data, our data sets are now made available on the open data site—www.data.gc.ca—as of July 1, 2011. We were one of the initial departments that participated in the pilot before even the launch of the open data site.

Currently, CIC has 37 tables on the open data site. These tables provide information about CIC's operations overseas and in Canada, the number of permanent and temporary residents admitted to Canada, and data also related to grants of citizenship to new Canadians. All data sets are made available as Microsoft Excel files, a common format, as you might know, useful for a large number of potential users and consistent with past updates. Thirty of the 37 data sets—very recently we were able to do this—are also being made available in machine-readable format, as comma-separated values, CSV, in order to offer users a version of the data that can be more easily incorporated into other applications. This CSV format, I understand, increases the interoperability of CIC's data sets.

[*Translation*]

The 37 data sets selected and posted by CIC are a mixture of tables that provide to Canadians and other people in other countries a broad background for the work CIC does and data on the number of temporary and permanent residents coming to Canada. These tables are a mix of those that are commonly requested of us by the public and those that we believe provide a useful overview of the work of the department.

[*English*]

Most ad hoc requests for CIC data are subject to the cost recovery requirements set out in section 314 of the Immigration and Refugee Protection Regulations. Other requests for data are received through the ATIP process. These are outside of the open data concept. By regularly updating and posting popular data sets on the open data site, CIC enables public access to these data easily, quickly, and at no charge.

Since CIC began making available its data sets on the open data portal, our data have been consistently among the most accessed on the site. This past March, five out of the top 10 data sets downloaded from the open data portal, and eight of the top 15, were from CIC. Of particular interest to the public were data sets on permanent resident application processing, in which we provide data on the number of applications approved and refused, and the processing times, etc.

The limitations l referred to earlier come from the personal nature, as you might have understood, of much of the data we make available. Of primary concern to us is that we protect people's privacy. While the data are ultimately derived from personal information from client applications, we only report aggregate numbers. Our challenge mainly is to make this data available in a manner that's as useful as possible, while also not releasing any information that can identify an individual.

● (0910)

[*Translation*]

As part of the Government of Canada's open data community, CIC representatives, namely, members of my team, regularly take part in open data working group meetings organized by our colleagues at the Treasury Board Secretariat, where we work with a number of other government departments to improve open data's visibility and usefulness.

[*English*]

The hackathon was mentioned already. We did participate very actively as well in that event, which I take as only the beginning of such events, with many more to come.

We see the open data initiative, to conclude, as being beneficial to the government, to Canadians, and to people around the world. For CIC, it allows us to distribute our data sets to a broad audience efficiently, and reduces the number of ad hoc requests that the department receives. For the public, it makes available data on citizenship, permanent and temporary resident processing, admissions to Canada, and information that has been shown to be both popular and useful.

[*Translation*]

Thank you for the time you have given me today.

[*English*]

Thank you.

[*Translation*]

The Chair: Thank you.

I will now give the floor to Ms. Montplaisir from the Department of Health.

Ms. Guylaine Montplaisir (Chief Information Officer, Corporate Services, Department of Health): Good morning and thank you, Mr. Chair.

I am Guylaine Montplaisir, the chief information officer for Health Canada.

It is a great privilege for me to be here today. I want to thank you for the opportunity to present the Health Canada approach and progress to date with open data.

I am pleased to report that Health Canada's open data efforts have led to the release of 71 data sets which are currently on data.gc.ca. This number fluctuates relatively dynamically as we clean the data sets submitted and add new ones.

Out of all government departments, Health Canada has the sixth highest number of data sets published on the open data portal. The majority of these data sets are related to drug products, natural health products, nutrient value of common foods, medical devices, adverse reactions and notices of compliance.

[*English*]

Health Canada is extremely pleased that some of our published data sets were used in the recent CODE event, the Canadian Open Data Experience appathon, and one of the applications that was placed among the finalists—it was called Munchables—was developed using Health Canada's data set on nutrient values and common foods. The application, if it were ever to be used, would actually enable Canadians to make better decisions about the food they eat and to make healthy food choices, which is one of our goals.

At the outset, Health Canada focused its efforts on publishing data sets that were already readily available on our other Health Canada sites. Our initial approach to identifying data sets for publication to the open data portal was to rely heavily on subject matter experts across our organization to come forward with data they wished to publish. Our approach evolved over time to the point that we're now actively seeking and soliciting data sets from the program areas, especially in high-value categories such as those that were identified in annex B of the G-8 Open Data Charter in June 2013.

We continue to reach out directly to program areas to identify additional data sets.

[*Translation*]

In summer 2013, Health Canada undertook the development of a vision document to guide future activities around open government, including open data. This effort intended to generate engagement and conversations about open government within the organization. This document outlines proposed approaches to finding data, such as soliciting input from stakeholders, evaluating web statistics to determine visits and searches of our Internet sites, and analysis of past access to information requests received. A full-year analysis of our web statistics placed the Canada Food Guide and the drug product database among the top searches.

Our vision clearly articulates Health Canada's commitment to open government. That means a commitment to foster greater health program transparency by the Government of Canada, the health regulator; provide Canadians with opportunities to participate in federal health policy development; steer innovation in health and life sciences; and ultimately encourage Canadians to make better informed decisions about their health.

[*English*]

The document also clearly enunciates a number of guiding principles, those most relevant to our open data agenda being openness, quality first, and stewardship.

On the openness front, we will strive to improve data and information sharing within and between organizations and cultivate a culture of "open by default" to dismantle silos and expand the data and information that is shared publicly.

On the quality first front, the data and other electronic information that we release to the Canadian public will be prioritized, easy to understand, and published in a convenient, machine-readable format that supports reuse. When we started publishing to the portal, when the first version of the portal was created, we were publishing in the format that the data existed in. Today, because we want to be open by default, we create the information in the format that is machine-readable. Today, 60% of our data is published in the format of CSV, comma-separated values.

On stewardship, the third principle, our plan will focus on building the long-term infrastructure and capacity to identify, manage, and make available the data that is solicited and captured by Health Canada on behalf of the Canadian public while fulfilling mandated responsibilities and activities.

In order to successfully implement the open government directive once it becomes effective this summer, Health Canada will establish a number of essential operational conditions, such as: putting in place an operational mechanism for the rigorous analysis of information and data that will respect to privacy, confidentiality, security and ownership before the data is placed in the public domain; maintaining an enterprise data set inventory, and we have already begun doing so; and establishing a process with the active participation of all program areas to help facilitate and prioritize their release. We will also recognize the need for sound identification and evaluation processes.

● (0915)

[*Translation*]

Going forward, Health Canada will continue to explore options to increase the availability of data on data.gc.ca. Work will be undertaken to facilitate the integration of Health Canada data with other sources such as energy projects data and weather data, which affect health.

This data will serve as the basis for applications that private industry can develop and make available to the public for use at home and on mobile devices. We will look to improve access to our data through the implementation of an application program interface for our more dynamic datasets, such as

[*English*]

recalls in safety—in good French—

[*Translation*]

to ensure Canadians have access to our most current data on an ongoing basis.

We will also continue to provide timely responses to stakeholder feedback with regard to the open data sets Health Canada has posted to the open data portal.

[*English*]

We will continue to identify data themes or clusters and prioritize their release, as part of the forthcoming open government action plan

2.0. Identifying and prioritizing data themes and clusters for public release will be based on two main principles: relevance to the Health Canada mandate and strategic outcomes; and responsiveness to what Canadians want and need to know.

Accordingly, analysis of the program alignment architecture and the strategic outcomes, as outlined in the Health Canada report on plans and priorities, will be the basis for categorizing information and data content. Stakeholder information needs will continue to be informed by environmental intelligence gathered from ongoing business operations, including stakeholder feedback, web metrics analysis, social media monitoring, as well as information release and analysis from our key international counterparts.

[*Translation*]

Mr. Chair, that concludes my opening remarks.

I appreciate the opportunity to be before the committee and am ready to address any questions you may have.

The Chair: Thank you. It is we who appreciate having you here.

I will now hand the floor over to Mr. Thivierge and Mr. Ram, of the Department of Transport.

Please go ahead.

Mr. Richard Thivierge (Director General, Business and Systems Architecture and Deputy Chief Information Officer, Department of Transport): Good morning, Mr. Chair and members of the committee.

[*English*]

I am pleased to appear before you today to discuss Transport Canada's open data practices and related contributions to the Government of Canada's open data portal.

With me today is Kash Ram, director general, motor vehicle safety directorate.

As you may know, Transport Canada was an early contributor to the open data portal. In 2011, the civil aviation aircraft registry data set was selected for publication as part of the portal pilot initiative. Today, Transport Canada has nine published data sets on the portal, which are freely available for download and use by all Canadians. These sets provide citizens with easily accessible, high-quality data that relates to a variety of Transport Canada programs.

Six of the data sets pertain to our motor vehicle safety program. They are the vehicle recalls database, vehicle recalls-last 60 days, the national collision database, the listing of vehicle manufacturers registered with Transport Canada, and both appendix F and appendix G of the pre-clearance list of recognized vehicle importers.

Two of the data sets pertain to the marine mode. They are the Canadian register of large vessels, and the Canadian register of small commercial vessels. One data set pertains to air transportation, which is the previously mentioned civil aviation aircraft registry.

In terms of the strategy pursued by Transport Canada for identifying potential candidate data sets to be published on the open data portal, given that the department has traditionally had a significant presence on the Internet, our strategy has naturally entailed leveraging information that is publicly available on our external website. We have also given priority consideration to data sets proposed by the Treasury Board Secretariat and by departmental business units.

In addition to public availability, other factors such as citizen and stakeholder demand for the information have been taken into account. For example, citizens had expressed an interest for data on subjects such as reportable motor vehicle collisions that occur on public roads in Canada, and this is among the reasons for our publication of several data sets relating to the motor vehicle safety program.

The department exercises due diligence prior to releasing data for public consumption. In each instance, consultations are held with the responsible business unit to determine the scope of the data and ensure that the data set can be released under the open government licence. All data sets considered for publication are then also closely reviewed for accuracy and to ensure that no sensitive content is present. As part of our rigorous multi-step validation process, business units are first required to verify the data and to certify that there is no security, privacy, or other restrictions that apply.

Our experience to date using the open data portal has been positive. Canadians have provided some feedback on the data that we have made available by submitting inquiries online via the portal. Although statistics are limited, our data has been viewed and downloaded hundreds of times and at the recent Canadian Open Data Experience national hackathon, Transport Canada was 15th for the most downloads of the 36 departments represented.

We also have noted that despite the prior availability of certain information, the publication of some data sets has resulted in a reduction in a number of public enquiries directed at the department in relation to the national collision database, and a simplified process for providing biweekly updates on vehicle recall data was devised and implemented.

Moving forward, Transport Canada intends to continue supporting the open data portal by actively assessing and publishing new data sets that provide quality data in relation to all areas of transportation that fall under the departmental mandate.

Thank you for your attention. We would be pleased to take any questions you may have.

● (0920)

[*Translation*]

The Chair: Thank you.

It is now over to Mr. Diverty, of the Canadian Institute for Health Information. Joining Mr. Diverty is Mr. Hunt.

The floor is yours.

[*English*]

Mr. Brent Diverty (Vice President, Programs, Canadian Institute for Health Information): Thank you. Good morning.

On behalf of the Canadian Institute for Health Information, I'd like to thank you for the opportunity to appear before the committee.

For the last 20 years, CIHI, as we're known, has played a unique role in Canada's health sector. As a government-funded but independent not-for-profit organization that provides essential information on our health system and the health of Canadians, our vision is simple: better data, better decisions, healthier Canadians.

Our mandate is to lead the development and maintenance of comprehensive and integrated health information that enables sound policy and effective health system management. Our strategic plan commits us to improving the comprehensiveness, quality, and availability of our data to support population health and health system decision making and to ensure its effective use. With our data access strategy, we ensure that our data is accessible to users through a number of ways in a timely manner.

CIHI is a data custodian for a wide variety of data on different aspects of the health system, including health services, quality of care, health expenditures, health care providers, and patient safety. Since our inception, and with input from our many stakeholders across the country, we have helped improve the depth and breadth of Canada's health data by developing information standards that allow every jurisdiction in the country to understand, compare, and use health data effectively; building and maintaining 28 pan-Canadian databases that enable jurisdictions to compare data; producing analyses on health and health care in Canada that are relevant, timely, and actionable; and increasing the understanding and use of data through education, reporting tools, and strategies.

Although we play an integral role in providing data and analyses to policy-makers in Canada's health system, we are neutral and objective in fulfilling our mandate. We neither create nor take positions on policy. We are funded by federal, provincial, and territorial governments and governed by a 15-member board of directors that links federal, provincial, and territorial governments with non-governmental health groups.

We work with a broad range of health organizations and partners across the country, providing data and information to help them fulfill their mandates. Our partners include provincial and territorial ministries of health that voluntarily submit data to us through data-sharing agreements, and other organizations such as Statistics Canada, the Public Health Agency of Canada, the Canadian Patient Safety Institute, Accreditation Canada, Canada Health Infoway, and Health Canada, with whom we collaborate on many health information initiatives.

Data quality and standards are fundamental to our work. Through our internationally recognized data quality program, we apply rigour to all data collection, analyses, and reporting activities. An integral part of data quality is developing and maintaining standards and working with our stakeholders. We take a lead role in developing and implementing national standards to ensure the consistency and accuracy of our data for use in pan-Canadian comparable reporting.

Much of CIHI's data is sensitive health information. As custodians of this data, we are committed to protecting the privacy of Canadians and take this role very seriously. We do this through a comprehensive information privacy and security program. As a prescribed entity under Ontario's Personal Health Information Protection Act, CIHI is one of only four organizations in Ontario authorized to collect, use, and disclose personal health information for the purposes outlined in the act. We are also subject to oversight by the Ontario Information and Privacy Commissioner, and our privacy practices and procedures are reviewed and approved every three years.

As we are all aware, open data generally is about turning the data of government, such as data on weather or climate, natural resources, processing times for immigration applications, things that my colleagues have mentioned, outward for others to use to improve transparency and to promote economic activity. Some data, such as the data CIHI holds, in many instances contains personal health information, and we need to safeguard that data and release it in ways that are appropriate. For example, we need to ensure that sufficient protections are in place to prevent re-identification of an individual, residual disclosure of their health conditions, or other sensitive information. Quite frankly, CIHI's existence depends on us holding up that expectation and requirement.

CIHI strives to make our rich public resource of administrative data from provincial and territorial health systems available and accessible to stakeholders in a way that ensures privacy and security. Just one of the ways we uphold those protections is by following a series of policies and procedures whenever someone requests access to a CIHI data set. We employ staff highly trained in data anonymization and other techniques for this purpose, and in some cases, hire other experts to advise us. This is how we were able to place two sample files of our acute care data set into post-secondary libraries in Canada last year.

● (0925)

Similar to the open data initiative, we have a multi-year strategy under way to make the data we hold more accessible to our stakeholders through a wide range of means such as OurHealthSystem.ca, which is a website designed to help Canadians understand how well their health system is performing. It features 15 indicators exploring five areas of performance that Canadians told us were most important to them.

Public reporting is a key part of our health system performance agenda. As part of this three-year package of work, we are developing a number of interactive tools featuring performance indicators for health regions, acute care hospitals, and long-term care homes. These tools will be publicly available on our website. The next release in the fall will provide cascading performance measurement reports for health regions and facility executives. Users will be able to export data on any of the 43 indicators available directly into Excel spreadsheets.

Quick stats, another of our products, is a series of free, static, and interactive data tables and supporting documentation about the health care system that is also available to the public through our website. They provide descriptive information for a range of purposes and are used by students, advocacy groups, media, and the broader public. The patient cost estimator, which looks at the cost of particular hospital procedures, and our "wait times for priority procedures" tool are interactive tools also freely available on our website that visually present complex data in a way that is easier for the user to understand.

Our analytical publications contain actual information on important topics for policy-makers, health care leaders, and the public. These products are enabled by the data we hold and the robust methodologies we use and maintain. Some of our recent publications focused on topics such as compromised wounds in Canada, drug use among seniors on public drug programs, adverse drug-related hospitalizations among seniors, bariatric surgery in Canada, measuring the level and determinants of health system efficiency in Canada, and end-of-life hospital care for cancer patients.

We also have a series of annual reports such as "National Health Expenditure Trends", "Regulated Nurses: Canadian Trends", some supply and payment information for physicians, and a comprehensive report on end-stage organ failure in Canada.

Finally, if the information someone is seeking is not available through any of the products mentioned so far, researchers, decision-makers, health managers, media, and the public can request information from one or more of CIHI's databases. Requests that are complex and require more work are completed on a cost-recovery basis.

In summary, CIHI operates in a manner that aligns well with what open data is trying to achieve for government. We make a considerable amount of data publicly available and have strategies in place to make even more data available in the future. Given our role as a custodian of sensitive health information, however, it's very important we strike the right balance between accessibility and the protection of personal health information. We continue to work with our federal, provincial, and territorial partners to achieve that.

I thank you for the opportunity to address the committee. Along with my colleague, Michael Hunt, I'd be pleased to answer any questions you may have.

● (0930)

[*Translation*]

The Chair: Thank you all for being here and for sharing your input with us this morning.

We are now going to turn the floor over to the committee members. We will spend about an hour on questions.

Mr. Ravignat has the floor for five minutes.

Mr. Mathieu Ravignat (Pontiac, NDP): Thank you, Mr. Chair.

I, too, want to thank you all for being with us today. There are a great many of you; I think this is the largest group of witnesses I have seen appearing before a committee, and I thank you.

[*English*]

I'll ask all of you to be brief because there are so many of you. My first question is a yes-or-no question. Have you ever suggested to Treasury Board that a data set be released on the open portal, and that request has been refused? If so, why?

[*Translation*]

The Chair: We will follow the same order we did for the presentations.

[*English*]

Mr. Bill Joyce: The answer from Statistics Canada is no.

Mr. Prashant Shukle: No.

Mr. Ümit Kiziltan: No.

[*Translation*]

Ms. Guylaine Montplaisir: The answer is no for me too.

Mr. Richard Thivierge: Same for us. The answer is no.

[*English*]

Mr. Brent Diverty: Our data is not on the open data portal.

Mr. Mathieu Ravignat: That's great, those are positive answers.

Has putting data sets on the open portal precluded you from putting data sets on your own websites or your own locations as you would normally do?

Mr. Bill Joyce: From our perspective, the data sets are discoverable—I used that word in my presentation—through data.gc.ca. The data themselves actually reside on the Statistics Canada web servers. They are discoverable, findable from the open data portal, but in that sense there is no duplication. They are discoverable from both locations.

Mr. Pierre Ferland: For us, as I said earlier, it's essentially an additional channel for publication. So it's better access for people.

Mr. Ümit Kiziltan: For us, we started publishing quarterly data before the open data came into being. Then when the open data became available as another venue, the current 37 tables that we have there, they are on open data. Five of them are also published on CIC's website. The same information is published at about the same time, and we don't see any problem or any conflict between those two venues at the same time.

● (0935)

Ms. Guylaine Montplaisir: For us at Health Canada, there are some data sets right now that are duplicated. They are typically the older ones. However, as we move forward towards renewing our websites, towards a healthy Canadian website, we will focus our effort and put the data on the data.gc.ca website and leave the other site to publish information.

Mr. Richard Thivierge: For Transport Canada, basically it is an additional channel. At this point we have not removed things or data sets that were available on our own website. So there is, to some extent, some duplication. It's an additional channel.

Mr. Brent Diverty: I don't think the question is applicable to us.

Mr. Mathieu Ravignat: This question is for Mr. Joyce. We heard Tuesday that there's quite an absence of data surrounding two groups on the open portal: aboriginal people and seniors.

I'd like to ask you why you think that might be the case.

Mr. Bill Joyce: Mr. Chair, I would like to thank the member for the question.

Statistics Canada does publish data on aboriginal peoples and on seniors, as well. These data are made available, made discoverable, through the open data portal.

I can't, from my perspective, address the wider question of data availability in that larger sense. My role within the agency is to think about the standard aggregate data that we have to publish for the broader community, and to make sure that those data are available in machine-readable formats.

One thing we will be doing in the future has to do with how, in some cases, data is wrapped up in publication format. It might be a PDF. It might be an HTML version of web language on the Statistics Canada website. Those data sources may not be downloadable. So there may be some cases where data is contained in publication format, which is not really a technically open format.

One of our next steps as an agency is to review those publications and make sure those data sources become open, machine readable, and downloadable.

[*Translation*]

The Chair: Thank you.

Thank you, Mr. Ravignat.

It is now Mr. Trottier's turn for five minutes.

Mr. Bernard Trottier (Etobicoke—Lakeshore, CPC): Thank you, Mr. Chair.

Thank you to the witnesses for joining us this morning.

[*English*]

I want to talk about this study that we've done and kind of give you a sense of why you're here today.

We started with customers—if you're ever trying to solve a business problem, that's a good place to start—and asked Canadians and representatives of different users of data what they'd be looking for. Then we did a scan around the world. What are other countries doing? What are other levels of government doing? Then we thought we'd end with our own government departments and get a sense of whether we are making good progress when it comes to open data.

What we've heard is that it's very important to have some raw data so that researchers and data experts can do things with the data, but also to have some synthesized data. Regular people on the street need to be able to access data also, and if there are some kinds of syntheses that different departments can do, that's very helpful. Also, because of all the data sets, hundreds or even thousands of data sets, there has to be some kind of search capability that regular lay-users can use to find out where that important data is.

There has also been mention of the billions or trillions of economic opportunity. A lot of that is actually savings in government, that actually opening data drives efficiencies in the sense of different people within government sharing data amongst themselves. If it's open data, you don't have to make expensive requests. It also makes it cheaper and more effective for Canadians to access data. It also drives really important benefits when it comes to decision making. Whether it's investments, or health, or safety information, it simply drives benefits. So there's unanimity that open data is a good idea. Nobody says open data is a bad idea.

My questions are more of a practical nature. We're trying to provide some recommendations to the Treasury Board, which is spearheading this initiative. There's a sense that these different open data initiatives are happening in each of your departments and all doing good work. Do you sense...? This is more of a polling question and maybe we'll go in the same order in which you made your presentations. Do you sense there's a need for more intervention from a centralized agency to do a horizontal initiative across your departments, whether it's Treasury Board or Shared Services Canada, if there were a role for a central coordinator of an open data initiative? Or are you better off doing it within each of your departmental verticals?

We want to get there. It's how can we get there with higher quality and more quickly. It's that kind of program management approach. What do you think would be the better way to achieve that result of trying to get to more open data?

●(0940)

Mr. Bill Joyce: Mr. Chair, from my perspective I would say that the options are not mutually exclusive. We, as an agency, have a duty to support the principles of open data in our ongoing publishing activities, but there is obviously value in a pan-governmental approach in which there is a certain coordination, where certain directives are in place. It holds us accountable and it allows a common approach. The end user of data may be using Statistics Canada data one day and data from Health Canada or Transport on another day, so obviously some consistency, from my perspective, is a good idea.

Mr. Bernard Trottier: Okay, thank you.

Mr. Prashant Shukle: From the geospatial data component, we were certainly working very closely with Treasury Board. I mentioned earlier the Federal Committee on Geomatics and Earth Observations, which is a government-wide approach to looking at the horizontal coordination of our geospatial data and information and which is, by and large, open. We are looking at business models that look at particular forms of efficiency. We're guided by the mantra of build once, use many times. In terms of our search capabilities, we're working closely with Treasury Board to also implement the mantra of search once and find everything.

We know now that technology allows us to do that. We also know that there is a collaborative will within the federal government to work together. I think that the strategic review process and the strategic and operating reviews have forced departments to think about how it is they work together, and as a result, the efficiencies and the strategic retargeting of how it is you do business has forced us to come together in more effective and efficient business models. I think the work of the 21 departments that have come together as the Federal Committee on Geomatics and Earth Observations is a critical example of that.

The work we've done at NRCan also speaks to a business model that we implemented back in 1999 with the provinces and territories. We have the Canadian Council on Geomatics. Canada has always had a leadership role in geomatics, so we work together collaboratively with 13 governments, provinces and territories, to share geospatial data in open and collaborative ways.

We have a GeoConnections program at NRCan. That program is responsible for international and national geospatial standards and ensuring that there is vendor neutrality, technological interoperability as well as data interoperability, and we've been given a cabinet mandate. I believe it was Minister Paradis that reannounced the program back in 2011, if I'm not mistaken. I can check that fact for you.

[*Translation*]

The Chair: Mr. Trottier, I am going to ask you to wrap it up to give the others a chance to respond.

[*English*]

Mr. Bernard Trottier: I want to get a sense of that horizontal versus a vertical approach to—

Mr. Prashant Shukle: We have both. We have horizontal and vertical. We're drilling down into the very specific component, and there is a horizontal approach. We're already working with Treasury Board on it.

[*Translation*]

The Chair: Does anyone else have anything to add?

Ms. Guylaine Montplaisir: On our end, we can really see the benefits of a horizontal approach.

[*English*]

Certainly from our perspective, a common approach and standards across the board are called for. As well, multi-level information-sharing agreements that would contain clear terms and conditions in line with the open data licensing agreement.... It would be a huge enabler from our perspective if the different health partners, whether they be provinces, territories, or other health partners that we deal with, would share the same rules. Then it would be easier to all fulfill our own role in terms of open data.

● (0945)

[*Translation*]

The Chair: Thank you.

Anyone else have anything to add?

Mr. Kiziltan, please go ahead.

[*English*]

Mr. Ümit Kiziltan: Very quickly.... The way we experienced this question in CIC is that we have the internal momentum driven by efficiency searches, trying to serve our clients better, and whatnot... and also driven by Government of Canada commitments, open government.

However, having said that, we also do benefit and we do appreciate the Treasury Board guidance, in terms of policy expectations, format, consistencies that aren't technical consistencies, and their support in terms of guiding us towards perhaps more preferred or more popular data requests.

I think that in that role, they, too, support each other very strongly.

[*Translation*]

The Chair: It is now Ms. Michaud's turn for five minutes.

Ms. Élaine Michaud (Portneuf—Jacques-Cartier, NDP): Thank you, Mr. Chair.

I also want to thank each of the witnesses for their presentation.

My questions are for Mr. Joyce and Mr. Béland.

This is my first time on this committee, and I think the study the committee has undertaken is very useful. I am glad to see so much focus on data accessibility and the principle of open data.

I had a quick look at the G8 Open Data Charter, as well as Canada's action plan for implementation and the principles set out by the charter. Beyond the technical side of accessibility, the charter indicates that quality information should be released in quantity and should be usable by all. I think that gives rise to a number of questions.

Let's consider, for example, what is happening at Statistics Canada right now. The government made a decision to get rid of the long form census. Just recently, Mr. Ferguson, the Auditor General, criticized the removal of data from the 2011 National Household Survey. In fact, 25% of the country's geographic areas were stripped of access to reliable data on their own communities. No one has even mentioned the potential impact on special groups who need that information. First nations and official language minority communities are two that come to mind. Some of my colleagues at the table today have no doubt repeatedly heard these same arguments in the Standing Committee on Official Languages.

How can your agency adhere to the principles set out in the government's action plan if you are stripped of the tools you need to provide reliable data to those who want and need it?

I'd like to hear your comments on that.

[*English*]

Mr. Bill Joyce: With respect, my role in Statistics Canada is to lead the dissemination program. I'm not able to address the specific issues relating to the exact nature or the funding questions relating to the specific statistical programs within our agency.

When we publish data from the variety of our statistical programs, including from the census and from the national household survey, we look at questions of data quality and it is my role to ensure that we present those data in the most open format possible. It's my role to promote greater access to the standard data.

[*Translation*]

Ms. Élaine Michaud: Beyond the technical dimension, Statistics Canada must have received feedback from users decrying the fact that the data is limited in its use.

Could you send us the feedback you have received in that regard?

[*English*]

Mr. Bill Joyce: I can note the question, but I am not able to respond to that type of question today. I do not have that information with me.

[*Translation*]

Ms. Élaine Michaud: Could you promptly send it to the committee?

I think every committee member would benefit from knowing how the lack of reliable data is really impacting users.

[*English*]

Mr. Bill Joyce: I will certainly note the question, and we will bring it back to the organization.

[*Translation*]

Ms. Élaine Michaud: Will you send the committee a written answer in a timely fashion?

Mr. Bill Joyce: Absolutely, I will do it as soon as possible.

Ms. Élaine Michaud: Thank you kindly.

Mr. Béland, do you have any additional information on the subject to share with the committee? Perhaps you could expand on Mr. Joyce's answer.

Mr. Yves Béland (Director General, Operations Branch, Statistics Canada): As Mr. Joyce mentioned, our area of expertise is the dissemination of data. Our mandate is to release the information provided by the various programs to the public. As Mr. Joyce said, we can take note of your questions and get the answers from the people at the agency who have that information.

● (0950)

Ms. Élaine Michaud: Thank you.

How much time do I have left, Mr. Chair?

The Chair: Thirty seconds.

Ms. Élaine Michaud: I am going to use that time to once again condemn the fact that certain types of data are no longer available.

This is a study that focuses on giving Canadians the broadest possible access to quality data. And yet, Statistics Canada is having to take down previously posted data from its Web site because it can no longer be used by those who need it. Quite frankly, I find it incomprehensible. It seems to go against the objectives that the government says it wants to achieve.

The Chair: Thank you, Ms. Michaud.

Mr. Adler, your turn for five minutes.

[*English*]

Mr. Mark Adler (York Centre, CPC): Thank you, Mr. Chair.

I thank all the witnesses for being here today.

I just want to begin with Mr. Shukle. You mentioned earlier that as more and more data sets become available through single portals, it becomes more difficult for the user to find that needle in the haystack.

Open data is all great, and by taking information and making it more readily available, just sort of pouring it out there, what is being done specifically to allow individuals to find the specific data sets that they are looking for?

There may be people, of course, who are adept at doing this kind of research, but most aren't. So open data is all fine, but for those lay-users who are just interested in finding specific information, who wouldn't typically go to a site like this but who are interested in finding out something specific, how user friendly is it for them? How are they able to be directed to the information that they need?

Mr. Prashant Shukle: I'll speak to my experience on geospatial data. I highlighted the fact that we had taken a very disciplined approach to our data, well over 100 years. You have to have a couple of things. One, you have to have really good metadata in a current context, and that could mean how you file and categorize your data. Then the data itself has to be managed in fairly standard formats and according to a rigorous process.

We actually have ISO processes, and we have international standards associated with how it is we manage our data. We actually play a leadership role in helping to define those standards so that we can get Canadian companies up at the forefront in being able to participate.

The second piece in the digital age with respect to that data is that you want to have vendor neutrality. You want to be able to have geospatial information, a map that can talk on an iPad should be able to talk on a Samsung, should be able to talk on a BlackBerry. You need that interoperability and you need standards associated within interoperability.

I guess I should say there are three things. The third thing is that I actually welcome the open data portal, because I think it condenses the number of federal portals. As a user myself, I find that finding the needle in the haystack is difficult because of the sheer volume of portals that exist and websites that exist. Having the open data portal forces a convergence and an ability for us, from a client perspective, to make sure that we find that data much more easily.

We're currently working with Treasury Board and others, like Google and the other big service providers, to sharpen the search engines. If you sharpen the search engines and you have better data, highly structured, available data, you should be able to search once and find everything.

That mantra that I talked about before, "search once and find everything", that's what we want to do.

Mr. Mark Adler: You mentioned about the ISO and the international standard. So Canada has played a key role in terms of making universal definitions. If you go to the Canadian single portal and you're looking for specific information, and then you go to the U.K.'s, the phraseology may be a little different. You're looking for the same kind of information, but they define it differently.

So there is this kind of international standard that Canada is playing a lead role in. Is this what you're saying?

Mr. Prashant Shukle: Again, I'll speak to the domain of the geospatial, which I would argue is the reference, because everything happens in a place, right? So, on that geospatial stuff, we work with the Ordnance Survey in Great Britain. We sit on the Open Geospatial Consortium with, I think, 437 organizations, a vast number of countries. We participate at the UN forum on global geographic information management. We're also doing the same thing from a satellite data perspective. So, yes, we're heavily involved.

We're helping to write the standards. We're helping to participate in the global discussions. We believe that Canadian firms and companies that deal in geospatial technologies should be able to participate at the international level because the volume of business opportunity is very significant, and it's by making sure that we play at that international level through an appropriate government role. But having a heavy private sector involvement and Canadian standards and leadership are key to an economic success.

● (0955)

Mr. Mark Adler: We're all aware of these data brokers out there, and all of that. Do you foresee a time when there will be so much of this public data out there that we'll need some kind of...? I guess the resources of the public sector are limited in terms of maybe a sense of all of that data. At some point, do you see the private sector being involved in the public data sphere of making sense of all of that public data and then being able to sell it to the private sector?

Mr. Prashant Shukle: I think the answer is yes. I think there's always going to be a role. I think that role is always going to evolve with the technologies. But I'll give you an example of a private sector company actually using the data. Kodiak Exploration, for example, reported to NRCan that $18 million in gold exploration resulted directly from open geological data accessed through our GeoGratis portal at NRCan. What they actually did is they combined the digital data from our topographic maps with other open online drilling data.

So you have either this brokering that's already happening in the private sector or you're probably going to get a rise of data management firms. I think that's what the big data analytics movement is all about.

Mr. Mark Adler: What was—

[*Translation*]

The Chair: Thank you, Mr. Adler.

[*English*]

Mr. Prashant Shukle: It's Kodiak Exploration.

[*Translation*]

The Chair: Since your time is up, I am now turning the floor over to Mr. Easter for five minutes.

[*English*]

Hon. Wayne Easter (Malpeque, Lib.): Thank you, Mr. Chair.

Thank you, everyone, for coming.

I did take a look at the CIHI data. I don't know why I've never seen it before, but it is impressive. In my area, one of the complaints we always get is that there are not enough doctors. We stack up fairly well in that area, but we're a little over on cost when I look at your data.

In any event, just to start off, I do want to add to Ms. Michaud's point. There's no need to get into a discussion on it, but there is increasing concern out there about the reliability of Stats Canada data, especially in the smaller communities. It was a government decision, and all I would say is that given the facts that have come forward, the government has the right to make the decisions they want, but I would hope the government reconsiders that decision, based on what we're seeing now as not reliable data in smaller areas.

I wanted to turn to the Health Canada information on your open data portal on, basically, information on drug products, natural health products, etc. That's a difficult area because I expect if you put up something that could be challenged by a drug company—negative impacts of a drug—you could face liability. Is there pressure from drug companies on what you put up? Do you find yourself in a position as a department maybe reluctant to put up what some of the side effects of a drug might be, given what I expect is a concern about liability?

Ms. Guylaine Montplaisir: We do routinely publish the information and the negative effects of drugs on our website through the Canada vigilance program. What we publish or don't publish isn't under pressure of industry or any other pharmaceutical organizations or anything like that. It's just information we routinely publish on our website.

● (1000)

Hon. Wayne Easter: One of the big concerns of people—and we actually have some experience in some of our constituency offices with this. The concern people have when they're put on a drug is what the side effects might be. How do you determine that? Do you determine it from the pharmaceutical companies' or the drug suppliers' information, or do you get that from doctors over time? How does that work?

Ms. Guylaine Montplaisir: The information is reported from multiple sources. It may be reported from people taking drugs. It may be reported from pharmacists or from medical doctors. It comes from all sources. It's aggregated and then it's published.

Hon. Wayne Easter: Thank you.

One of the other areas I found rather interesting in your presentation was that you seem to be moving to provide energy projects data as well as weather data. Why is that?

Ms. Guylaine Montplaisir: I simply talked about the need to look at sources of data more horizontally. We know that things like the weather have impact on the health of Canadians. Whether you live up north or in other regions of the country, we see different patterns from a health perspective. So we see the value in integrating data and we really want to move in that direction.

Hon. Wayne Easter: We always find you're in a better mood if you're living in P.E.I.

On natural resources, you partly, I think, answered this question of Mr. Adler's. You related it to some mining, I guess, in Canada, but how do you see this information from open data being utilized more to provide economic opportunities for some of the private businesses out there? I can certainly see it both in the forestry area and in the mining area, but could you give us other examples?

[*Translation*]

The Chair: Kindly keep it brief, as there's only a minute left.

Mr. Pierre Ferland: First of all, it serves the principle of transparency. Ensuring a certain degree of transparency is important, especially when it comes to the mining and extraction sector. It's important to inform Canadians about extraction activities across the country.

Now, as for forestry.

[*English*]

The land mass analysis and what we publish in terms of the—there are technical terms related to this. They really provide companies the opportunity to decide what their business strategy is going to be by way of exploitation in the future. We think those assets are essential and useful for businesses and people as well.

[*Translation*]

The Chair: Thank you.

Mr. Easter, you're out of time.

Ms. Ablonczy, please go ahead for five minutes.

[*English*]

Hon. Diane Ablonczy (Calgary—Nose Hill, CPC): The CIHR has a lot of data available, but it's not using the open data portal that your sister organization in the U.K., for example, does. I'm wondering why not and if you're moving in that direction.

Mr. Brent Diverty: The data we hold is provincial and territorial data that comes to us, so we're the custodian of the data. Historically, we've not participated in the open data portal because we're not a federal government department or agency.

Hon. Diane Ablonczy: Neither are municipalities or provinces, but they're all cooperating.

Mr. Brent Diverty: We're pleased and interested in cooperating in the portal. It's just that, to date, the way we worked with our jurisdictional partners, we haven't begun to do that yet.

Hon. Diane Ablonczy: Is there an openness to doing that? I'm not trying to give you a hard time, but as we're trying to direct people to this portal and have a very holistic range of data, if something as important as your organization isn't participating, then people don't have an important piece of the puzzle.

Mr. Brent Diverty: Absolutely, and not unlike Statistics Canada, I think there's an opportunity to have some sort of a redirect towards the data that's on our website from the portal. I think that would be a good way to perhaps do this efficiently.

● (1005)

Hon. Diane Ablonczy: So is the answer no, you're not open to it?

Mr. Brent Diverty: I think the answer is yes. We are open to it.

Hon. Diane Ablonczy: But only by link....

Mr. Brent Diverty: I'd be happy to explore further what the best means of participating in the portal would be. We are open to participating in the portal.

Hon. Diane Ablonczy: Okay.

Back to Health Canada, Ms. Montplaisir, you mentioned that you're actively seeking out and soliciting data from areas that have been identified as high value. I wonder if you could expand briefly on that, because I have some other questions.

Ms. Guylaine Montplaisir: The areas of high value were identified, as you know, on the G-8 annex, the G-8 open data agreement annex. So as a result, when these got published, we actually sought out the different program areas within our organization to find out which data sets they had available and what was in a format that could be readily published.

We did find quite a few areas where we've been facing challenges with respect to the format that the data is in and we are continuing to work with these program areas to find technological ways to publish this information as rapidly as Health Canada can.

Hon. Diane Ablonczy: Can you give us a quick example?

Ms. Guylaine Montplaisir: An example would be with respect to the medical device active licence listing, for instance.

It's an area that is, however, highly dynamic and we believe that it could be better published through an API, an application programming interface, because publishing any of this information in any different way, we would be putting stale data out there and it wouldn't be useful on an ongoing basis.

Hon. Diane Ablonczy: Okay, that helps a bit.

CIC talks about the work that's being done to make geo-data available through the open portal and I'm wondering if you have a timeline on that. What's the ETA for that addition?

Mr. Ümit Kiziltan: As I explained, the source data is personal data—

Hon. Diane Ablonczy: Sorry, I asked you the wrong question. That was really for NRCan, I apologize.

Mr. Prashant Shukle: By default, our data is open, so it's already on the open data portal. We also have geo-based data sets—federal, provincial and territorial—that are available through the geo-based portal. We have annual agreements with the provinces to update them, so they get automatically updated annually in terms of the data sets. We're looking at making new data sets available over the course of the year and I can certainly endeavour to provide information to the committee around when we make new data sets available.

Hon. Diane Ablonczy: I just refer to your—

The Chair: Thank you, Madam Ablonczy.

[*Translation*]

Your time is up.

Mr. Ravignat, your turn.

Mr. Mathieu Ravignat: Thank you, Mr. Chair.

Making data more accessible can result in costs, especially in terms of human resource requirements.

In light of the current fiscal situation, I'd like you to describe the financial challenges you face in making data accessible and the impact they have on your ability to do just that.

With your permission, Mr. Chair, I will ask Mr. Joyce to answer first.

[*English*]

Mr. Bill Joyce: The raison d'être for our agency is to publish the data and make it available. So we see it as a very good match to the mandate of our agency. There has been a long history already toward making data available in machine-readable formats and in open formats. Our publication processes for aggregate data for statistics are geared toward that end, so there is not an additional substantial cost in providing machine-readable open formats of our data.

Mr. Mathieu Ravignat: You're already in that business.

●(1010)

[*Translation*]

Mr. Bill Joyce: Yes, indeed.

Mr. Pierre Ferland: Your question fits into the broader context of information management, which is a real challenge for the government and the private sector.

The question is whether we can organize the data in such a way that it can be manipulated mechanically and automatically, without the need for human intervention or an additional investment to cover costs. Over the past 30 or 40 years, information technology has evolved fairly organically, and now we're at the point where we have mass quantities of information and data. I would qualify our ability to publish that data in an open format as really a subset of the broader challenge of managing information. It really reflects the strategic direction set by the government. Information management policies will really move us to a place where we can release information to the public at a lower cost.

I don't know whether that answers your question.

Mr. Mathieu Ravignat: As things stand, you don't face any challenges when it comes to resource requirements?

Mr. Pierre Ferland: The real challenge lies in how the information is organized, stored and produced. Once that aspect is resolved, resources will, by extension, pose much less of a challenge.

Mr. Mathieu Ravignat: Things will be harder initially, but run smoothly eventually.

Mr. Pierre Ferland: Precisely.

Mr. Mathieu Ravignat: Very good.

Mr. Pierre Ferland: It's a matter of how the data is organized.

Mr. Mathieu Ravignat: Thank you.

[*English*]

Mr. Ümit Kiziltan: In our experience, open data did help in terms of the efficiency of our work, so we're getting less ATIP, less cost-recovery. Those are helping us save resources and invest those resources in making more data available, not sending out DVDs or mail-outs and whatnot.

Having said that, we don't experience any pressing challenges on the resources, but if one had more resources, one always asks if one could do more. My answer to that would be yes, but we don't face challenges that are preventing us today to respect the principles and the current operations of open data.

[*Translation*]

Ms. Guylaine Montplaisir: I agree with the points my colleagues made, so I won't repeat them.

However, I will say we definitely have a challenge as far as data publication and official languages are concerned.

Internally, the datasets contain text that our researchers and scientists use and sometimes that text is, by default, accessible in only one language. So publishing a dataset always poses that challenge. There is a requirement to release in both official languages metadata that, under normal circumstances, is available internally in just one language. They may not be automatically recorded in the same way, so that has to be done. But I think the key lies in avoiding the duplication of work and opting for an open approach right from the get-go. That means taking the steps that need to be taken and putting the necessary processes in place to do the work the right way and make the data accessible to Canadians in the appropriate format.

Mr. Richard Thivierge: Mr. Ravignat, thank you for the question.

I agree with the previous witnesses.

Clearly, the way the work is organized is an important dimension. Earlier someone mentioned that the key was to take advantage of a new distribution network, and that means no longer doing some of the things we did in the past. As a result, we have been able to redirect information dissemination resources to the new activities. For instance, a great many documents, databases and publications were used to make the data available, but it was paper-based or kept on a CD. Staff had to spend a lot of time and effort to produce and distribute that data.

In today's climate, Canadians are increasingly equipped to handle the self-serve method. But it does present some challenges, as pointed out during the discussion earlier. The adoption of standards is a critical element in ensuring that the consumers of today's data age—businesses and individuals—are able to find their way. Certainly, challenges exist as far as organizing the information goes, but resources have been shifted to new distribution methods.

The Chair: Thank you.

Did you have a comment to wrap up the discussion, Mr. Diverty?

[*English*]

Mr. Brent Diverty: Just very briefly, like Statistics Canada, it's part of our mandate to make sure our data is accessible. In fact our success is largely based on the extent to which we make data available to the range of users. Obviously, there's always more needs to satisfy than you can satisfy, and you have to make choices around that. But historically we've spent a lot of money assembling data sets, and as those data sets have matured, now we're turning more of our resources to access and supporting users in using the data through understanding and use.

● (1015)

[*Translation*]

The Chair: Thank you.

Mr. Aspin, you may go ahead for five minutes.

[*English*]

Mr. Jay Aspin (Nipissing—Timiskaming, CPC): Thanks, Mr. Chair.

Welcome to our guests this morning. Thank you for helping us with this study.

I have a couple of questions. My first question is to Mr. Shukle of NRCan. Somebody suggested that possibly Natural Resources would be a more logical home for the location of the open data portal, given that the majority of the data sets belong to this department, and of course, your longevity in this space, since 1842, I believe you said, when you started. Do you agree with this suggestion?

Mr. Prashant Shukle: Our business is really in the geospatial world. I think open data is a broader definition. We're certainly happy with the way the situation is currently. We're happy sticking to our knitting.

Mr. Jay Aspin: Would you also be happy to share best practices with other departments?

Mr. Prashant Shukle: Absolutely.

Mr. Jay Aspin: My second question would be to CIC and Health. Recognizing that the data in both of these departments is personal information, sensitive information, confidential information, how do you anonymize this data? How do you protect the confidentiality?

I guess I could start with you, Ms. Montplaisir.

Ms. Guylaine Montplaisir: The data we publish is only aggregated data. We don't publish any individual-level data on the sites, so right away we don't have to really deal with the anonymization aspects of it. But of course privacy, security, and confidentiality are always of utmost concern to us. As the information management lead in our organization does the approbation of the release of data sets, we make sure that all of these concerns are addressed right from the get-go.

Mr. Jay Aspin: Right from the get-go, you sort it out right from the beginning.

Ms. Guylaine Montplaisir: Absolutely.

Mr. Jay Aspin: From CIC, Mr. Kiziltan....

Mr. Ümit Kiziltan: We're very similar. In our case we also cannot make our source data directly available as is. So we aggregate our reports, our tables. Data sets are all aggregate data.

Regardless, even if it's aggregate, we need to deal with the smaller values to prevent identification of individuals, so we sometimes group smaller values. We mask them. In a sense, we drop certain values. We also have algorithms to randomly round data. As we round them, it helps to anonymize certain instances and also systematically at a different layer. Within the department we consult with ATIP colleagues to ensure that, again, we are respecting privacy.

This, of course, poses a challenge and puts a limitation on the type of response to different requests. Let's say we have received a very recent request through the TBS portal, where the clients can come and suggest new data sets they would like to see. We have separate data sets, for instance, on immigration category admissions to Canada. We also have source-country separate information. But the request, for instance, is to cross them, so when we cross them, we know there will be very small cells. As the challenge comes, we work on it, and that obviously delays a little bit our making this data available. What you're pointing at is definitely an everyday question for us.

Mr. Jay Aspin: Thank you, sir.

Mr. Chair, do I have any time?

● (1020)

[*Translation*]

The Chair: Very little.

[*English*]

You have 30 seconds.

Mr. Jay Aspin: That's good.

Thanks, Mr. Chair.

[*Translation*]

The Chair: Ms. Michaud, your turn.

Ms. Élaine Michaud: Thank you, Mr. Chair.

My questions are for the Statistics Canada officials.

I think these may be more in line with your area of expertise, so you may know the answers.

Some witnesses suggested that it was important to have additional data at a more granular level, such as local data or data by industry. They would like public use microdata files, which are not currently available on the federal government's open data portal. The files are, however, available upon request from Statistics Canada free of charge.

Why are the public use microdata files not available on the federal government's open data portal or in the CANSIM database?

Mr. Bill Joyce: Yes, absolutely. Thank you for the question.

[*English*]

Statistics Canada does have programs to allow access to researchers, to allow access to microdata files. In all cases, even with what we call public use microdata files, additional licensing restrictions apply. For this reason we can't apply a true open licence. We do apply some of the principles of open data. For example, these data files are made available free of charge.

The extra licensing restrictions prevent the merging of files. You've heard a lot about the open data community, about the concept of data mash-ups and merging and linking data sources. That's the one thing that we can't allow with our microdata files in order to protect the confidentiality of Canadians. So it's a question of the additional licensing restrictions that prevent us from considering these files as truly open.

[*Translation*]

Ms. Élaine Michaud: So licensing restrictions are really why someone would have to request them from Statistics Canada.

Now I have a question for the Transport Canada officials on data availability.

Do you have information about transport safety when it comes to airlines, trains and cars? If so, do you release the information on the federal open data portal?

Mr. Richard Thivierge: Thank you for the question, Ms. Michaud.

As I explained earlier, we constantly assess what can be released on the portal. That said, the data does exist.

Transport Canada puts out an annual report on transport safety. It lists, in chronological order, all the incidents and accidents for each method of transportation, as well as related statistics. We are in the midst of figuring out how we can release that chronological information on the portal.

Ms. Élaine Michaud: Is this data currently available only upon request? How accessible is the data to the public or those interested in it? One year after the Lac-Mégantic events, people have a lot of questions. Interest in that type of data must have increased. Is the data available? What criteria do you use to decide which data will be made public and which will not?

Mr. Richard Thivierge: Traditionally, Transport Canada has followed Statistics Canada's best practices—in other words, rules on confidentiality. A lot of information is assessed based on confidentiality and other criteria.

To answer your question more directly, I believe the last annual report on transportation was produced back in 1996. More recent data is available online. That annual report contains more than 200 tables and time series, and it is available on Transport Canada's website.

The Chair: Thank you, Ms. Michaud. Your time is up.

I will use my prerogative to ask a question. This is the last meeting scheduled to hear testimony related to our study.

My questions will be addressed primarily to the representatives of Health Canada and Transport Canada, since these issues concern them more. I would like to talk about the idea of having a portal that would pool data from various levels of government.

Different levels of government have departments that are very similar to Health Canada and Transport Canada. What do you think about the idea of setting up a portal that would bring together data from various levels of government? A number of witnesses have told us they would like to have access to data from different levels of government in order to cross-check it and carry out research that is more specific than that based on exclusively federal, provincial and municipal data.

●(1025)

[*English*]

Ms. Guylaine Montplaisir: From a Health Canada perspective, most of the data we consume comes from either our partners or the different levels of government. As a health regulator, we produce a certain level of data and the other data is published directly by the provinces or through our partner at the table, through CIHI. We access data from the provinces through the data that we get through information-sharing agreements from CIHI, or directly from the provinces and territories.

Through the open data portal today, you can get access to the provincial and some municipal sites. Seeing them on a same site, I would see some definite value to Canadians, because today they probably struggle with the search engines trying to locate all of the information they're looking for. The more we move to a portal that is more integrated, the better off we're going to be. As a federal government, though, I can see this as a significant challenge with the provinces.

[*Translation*]

The Chair: Thank you.

Mr. Thivierge, go ahead.

Mr. Richard Thivierge: Mr. Chair, thank you for your question.

The federal government does have a presence and a role to play, more in terms of certain modes than others. As we know, in Canada, the highway system mostly falls under provincial and municipal jurisdictions.

That said, we are working with the provinces in a variety of areas to get an overall view, across Canada, of the four main modes of transportation—rail, sea, air and road. The idea of a common portal seems to be very relevant and is likely to provide major benefits.

That said, the project nevertheless presents some significant challenges. Similar practices would have to be established for the management of that website, be it in terms of data renewal frequency, adoption of standards or adoption of an identical taxonomy—whereby the same terminology would be used, either in terms of secondary or connecting roadways. All those challenges must be taken into account before we head down that road.

As I mentioned, we are already collaborating very closely on certain files. We are producing joint reports with provincial and territorial governments, including a report on the national highway network. That report is the result of a federal-provincial collaboration. The information has been posted on the website of the Council of Ministers Responsible for Transportation and Highway Safety and the Council of Deputy Ministers Responsible for Transportation and Highway Safety. This could provide an idea of what types of products could be made available on a common portal.

The Chair: Thank you.

Mr. Shukle, I see that you want to add something on this topic.

[*English*]

Mr. Prashant Shukle: Basically, I only wanted to add that the road network that Richard is talking about, we also have that road network, which is negotiated with provinces and territories, and we actually get that. We're working with municipalities as well, currently 200 municipalities, to update that road network. That road network is one of the layers that we provide as part of the geospatial data, and it also includes rail networks.

[*Translation*]

The Chair: Okay.

Mr. Ferland, I ask that you be brief, since time is running out.

Mr. Pierre Ferland: Mr. Chair, the question you asked goes to the core of the open data initiative. Here is what we need to decide: will we be the institution that gathers all the data or the institution that makes this data available?

I would like to come back to a question asked earlier about the private sector's participation. In the case of significant demand, we expect the ecosystem outside our walls to understand that demand and pool the public and private information in order to build a consistent and cohesive product that meets the demand. The question is whether this will be done by the government, the industry or external stakeholders who will seize the opportunity and create the product for the benefit of Canadians and companies.

●(1030)

The Chair: Exactly.

Thank you for your answers.

This concludes today's meeting in terms of all the testimony related to our study.

I want to thank the witnesses for being the last ones to come meet with us to share their departments' and agencies' expertise.

I invite the committee members to remain here for another 15 minutes to discuss committee business.

I will suspend the sitting for a few minutes.

[*Proceedings continue in camera*]