



HOUSE OF COMMONS  
CHAMBRE DES COMMUNES  
CANADA

# **Standing Committee on Government Operations and Estimates**

---

OGGO • NUMBER 022 • 2nd SESSION • 41st PARLIAMENT

---

**EVIDENCE**

**Thursday, May 1, 2014**

—  
**Chair**

**Mr. Pierre-Luc Dusseault**



## Standing Committee on Government Operations and Estimates

Thursday, May 1, 2014

• (0845)

[Translation]

**The Chair (Mr. Pierre-Luc Dusseault (Sherbrooke, NDP)):** We will begin our 22nd meeting.

Today, we are hearing from two witnesses by videoconference. We have a connection with Michael Chui, partner at the McKinsey Global Institute, joining us live from Miami, the United States. Afterwards, we will establish a connection with the other witness, Mr. Baker, the Chief Executive Officer of the Chicago Open Data Institute.

As usual, we will let our witnesses make their presentation for our study on open data. Afterwards, the committee members will have an opportunity to put questions to the individual of their choice.

Mr. Chui, thank you for joining us.

[English]

You will have 10 minutes for your presentation.

[Translation]

Go ahead.

[English]

**Mr. Michael Chui (Partner, McKinsey Global Institute, McKinsey and Company):** Thank you for the honour of being able to spend some time with you. Even though I'm in Miami now and I live in San Francisco, I actually was privileged enough to have grown up in Burlington, Ontario. It is truly an honour to be able to interact with this committee. I did prepare a few brief remarks, which I'm happy to share with you, but I'm actually looking forward to the conversation.

As introduced, my name is Michael Chui. I'm a partner at the McKinsey Global Institute, which is McKinsey and Company's research arm. I lead some of our firm's research on the impact of long-term technology trends. Basically I'd like to share with you a few of the findings from some of the research we conducted.

We published a report in October entitled "Open data: Unlocking innovation and performance with more liquid information". Clearly, as I think people on the panel are aware, open data has become an increasingly important trend around the world, with over 40 countries having implemented open data portals. While a lot has been written about the importance of open data to unlock transparency as well as accountability in government and public institutions, we really focused on the economic potential that could be unlocked using open data.

Just to explain what we meant when we did our research, we actually viewed open data as being defined or varying across four dimensions.

The first was accessibility, or simply the number of people or the number of entities with access to data. Where more people had access to data, we considered it to be more open.

Second, we also considered machine readability. Of course, almost all data in some form can be machine readable, but some forms are easier to use, easier to process, such as comma-delimited and other formats. That was another dimension that we considered to be important.

Third, we also considered cost. When information is made less expensive, or is free, it's more open. Again, sometimes governments and other institutions implement some sort of cost recovery. We didn't want to say that data was completely closed if a modicum of charge was associated with it.

Finally, the fourth dimension we described involved the rights to use that data, whether it could be redistributed, how it could be processed, etc. Data could be completely unencumbered in terms of legal rights to use, or there could be some restrictions on it. We think that varies along the continuum. We really think that data can be more closed or more open or more liquid, as we described it, rather than just open data and then everything else.

That being said, what did we find when we looked at the potential economic impact of open data? We looked across seven different sectors of the economy. The sectors include education, transportation, consumer products, electricity, oil and gas, health care, and then various aspects of consumer finance. When we looked across all of those different sectors of the economy and we looked globally, we found that an additional \$3 trillion to \$5 trillion in impact could be created using open data. These benefits include increasing efficiency, developing new products and services, and even consumer surplus, which is the type of benefit that individual citizens can obtain when they have access to more open data or to applications that use open data.

There are a few other findings. Open data also enhances the impact that big data can produce, which has been another area of study for us. Oftentimes, when you combine data from multiple sources, you can actually derive more value. Some of the ways in which you derive value include increasing transparency, exposing variability, enabling the ability to conduct experiments in the real world, segmenting populations to tailor actions, augmenting or automating human decision-making, and then defining new products and services. Really when we looked across the board, if you think about exposing variability and enabling experimentation, about one-third of all the impact we found came from the ability to benchmark, to compare yourself against others.

We also found that individual citizens stand to gain the most from open data. Over half of the impact we found—again, that's not separate from benchmarking, because you can do individual benchmarking as well—in terms of potential benefits would actually accrue to individual citizens or consumers. We found in fact a very closely related concept to open data, which we described as “my data”. That's where an individual citizen or person has access to data that a government or a company has about them. That was one of the sources of benefits that individuals could have, for instance, my ability to compare my health care outcomes with people who are similar to me.

● (0850)

Open data can also help businesses raise their productivity and create new products and services. Companies clearly benefit from the ability to benchmark both internally as well as externally. Open data can also be used to create more tailored products by providing more consumer insights. Of course, open data also creates new risks around reputation and potential loss of control over confidential information, whether it be personal information or corporate or organizational information.

We also think that governments have a truly central role to play as a source of open data, which clearly a number of governments have been leading in that, as a catalyst for the use of open data, as a user itself of open data, and also as a policy-maker. Clearly, government has a tremendous amount of data that it could make available, and increasingly does.

The other interesting thing is if you go back to the point that I just made, which is that a lot of the benefits actually can accrue to a diffuse set of consumers or individual citizens, if you believe that's true, then in fact government is one of the entities that has the potential to actually speak for that diffuse set of groups rather than any special interest group and thereby implement policies that make the benefits of open data more likely to be captured.

The last point I'd make is this. While making data more liquid, making it more open often is an unnecessary action in order to capture some of this value and it's often not sufficient. Other things that have to happen are that you need to create a vibrant system or ecosystem of developers who actually use the data to create applications, because most people won't look at the raw data itself; they'll use applications that take advantage of the data. Open data, as a result, often has to be combined with other sources of data. You need thoughtful policies around intellectual property, privacy, and confidentiality. You'll need to invest in technology along with

investment and skills. This is clearly one of those areas where we found a tremendous gap between the need for these skills and the actual supply of them.

Standards also have to be developed in order to make data comparable from multiple sources. Then actually releasing metadata, data about data, can make open data more usable.

In closing, the potential benefits of open data truly can be transformative—as we said, it's in the order of trillions of dollars annually on a global basis—but they can often be self-reinforcing. When open data is made available and applications that are useful are actually developed based on the open data, that often encourages more open data to be released and then that cycle continues.

Let me conclude with that. Hopefully that was a helpful tour of some of the research that we've conducted on open data.

**The Chair:** Thank you very much for this presentation.

[*Translation*]

Since our next witness has not yet appeared, we will begin the question and answer period right away.

Mr. Ravnignat, you have five minutes.

**Mr. Mathieu Ravnignat (Pontiac, NDP):** Thank you for being here.

I would like to hear your thoughts on a few issues.

How do you think the government could ensure that the data it transfers or makes public is adapted for commercial use? As you know, that is one of the Canadian government's objectives. Some witnesses have told us that the data was not quite useful. However, this is the first opportunity we have to talk about the data's commercial aspect. So I would like to know whether you have an opinion on how to go about this.

Thank you.

● (0855)

[*English*]

**Mr. Michael Chui:** There are a few things that we've learned from working with private sector clients. You can think about what you need to do with open data as almost being like marketing and the aspects of marketing which we think are applicable.

Number one is understanding the need that's out there. Just as a marketer wants to understand what product or service to deliver by deriving customer insight it's helpful for the government in this case to understand what the needs are for data out there. Ways to find out about that include convening groups, customer focus groups in this case. It would be commercial focus groups where you would actually ask companies what data would be most valuable to them. In fact, at our firm we've actually brought together groups of business executives and made them aware of some of the open data that was available from governments. They were actually surprised and said that this is something they can actually use. Being able to create that awareness which is the first part of marketing is incredibly important.

Then there's continuing to understand... Right now open data efforts are often supply driven. You know, there's a bunch of data and we'll just throw it out there. Again, you need that demand signal to come back through an open dialogue with your "customers".

**Mr. Mathieu Ravignat:** It's fair to say that one of the essential pillars of making data available is that dialogue piece. It's making sure that you connect with stakeholders at various levels to ensure the data is actually useful. It's something that unfortunately we've seen very little of in the open government strategy of this particular government, but that's neither here nor there.

The other question I wanted to ask you was whether or not you've had a chance to compare some of the efforts done by the Obama administration with the Canadian efforts, and whether or not you have an opinion with regard to the quality of both initiatives.

**Mr. Michael Chui:** I really haven't had an opportunity to compare the two initiatives. I don't think I could provide an informed perspective on that.

**Mr. Mathieu Ravignat:** Thank you.

Coming back to the commercial use of data, one of the issues is making sure it's a fair level playing field, that those companies that perhaps get procurement contracts with the federal government don't get some kind of favouritism with regard to access to certain data. You obviously don't want to advantage one particular sector over another.

I wonder if you have any thoughts with regard to ensuring the fairness of the access to data.

**Mr. Michael Chui:** I'm not an expert on procurement, so I'm not sure I could comment on that, but certainly when data is made available broadly through a portal, or what have you, in general, anyone who has access to the Internet or the web is able to access the data. Certainly from a pure access standpoint, I think that type of equality of access can be ensured.

**Mr. Mathieu Ravignat:** Thank you.

[*Translation*]

**The Chair:** Thank you, Mr. Ravignat.

Mr. Baker just joined us live from Chicago. I think he can hear us.  
[*English*]

**Mr. Paul Baker (Chief Executive Officer, Chicago Open Data Institute):** It's Paul Baker here. I hear you.

**The Chair:** Good. We will start with your presentation.

**Mr. Paul Baker:** Sorry I was late. A water main flooded the freeway in Chicago and all the traffic had to leave. It took an hour and 15 minutes for about a 20-minute trip.

• (0900)

**The Chair:** It's all right. We just heard from Michael Chui, from McKinsey and Company, from Miami. So now we will hear you, Mr. Baker, from Chicago.

You have 10 minutes, and then we will have questions from the members here at this committee.

**Mr. Paul Baker:** I was sent a few questions, or things you'd like to know about. I have a bunch of notes about those. I don't know if you're going to ask questions related to what you sent.

As for my background, I've been active in the open data movement in Chicago for about six or seven years. Chicago has gained a reputation as the open data capital of the United States, and even when I travel internationally, people seem to know about Chicago's open data efforts. Government is responsible for some of that. There are a lot of independent designers and developers who've been pushing for open data, lawyers pushing for open data, organizations like Common Cause.

In the United States, the open data movement was initially about people looking for political transparency, wanting to know who was making political contributions. Much of the initial impetus was because, during the Nixon period, the Watergate burglaries were financed by secret corporate donations. The Nixon campaign had caused Common Cause to be formed as a bipartisan Republican and Democratic group that was in favour of transparency in political donations.

From that, government started collecting data. Computers got a lot more powerful. Data was released. The issue of open data is now not only political transparency, it's also efficiency. It's the idea of government as a platform. Whether it's much government data, open data can be used to create businesses like Google Maps or weather reports. Some companies are aiding farmers trying to figure out when they should plant a crop, when it's going to rain, when it's not going to rain, whether they should irrigate, and these kinds of things. There's a lot of efficiency and economic benefits to open data that have come to the fore in the last three or four years as a lot more local and national data has been released.

A couple of weeks ago, the federal government released a bunch of Medicaid paid claim data. There are several other sets of data they're going to release, several other sets of data that have been released. This and other types of data, electronic medical record data, is going to be available to doctors and hospitals and clinics treating Medicaid patients. Much of it's already available. According to the Affordable Care Act, or Obamacare, patients and doctors treating Medicaid patients are required to be able to receive electronic medical records. That's going to change the whole way people are treated.

Genomic data now is being combined with electronic medical record data to do medical studies without actually having to devise an experiment and do blind tests with control groups and that type of thing. You can just look in the data and look for patterns in that data. So, maybe in women treated for breast cancer, some live and some die. You look at the ones who have lived and you look back through how they were treated. You look at their genomic structure, and you look for particular medicines that can treat different types of diseases based on genetic traits and particular drug regimens.

That's a very broad view of what's going on in open data at the federal level. We could maybe get into the city level and the state level a little bit later.

[*Translation*]

**The Chair:** Thank you for your presentation.

We have already begun the question and answer period. We will continue with Mr. Trottier for five minutes.

I ask that you specify to whom you are putting your questions.

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

[*English*]

Thank you, guests, for being here this morning.

Mr. Chui, I want to ask you some questions about the McKinsey report published in October 2013. You mentioned in your remarks that you focused on seven sectors. In the report, you identified an economic value of \$3.2 trillion to \$5.4 trillion annually worldwide just in those seven sectors.

How is it that McKinsey chose to focus on those seven sectors as opposed to others? Obviously, there are some other big sectors in a country like Canada: agriculture, fisheries, mining, and even tourism. I suppose the overall economic value could be much greater when you start considering those other sectors.

Also, if you look at only the United States, in the report you identified a value of \$1.1 trillion. Given that the size of the American economy vis-à-vis the Canadian economy is about 11:1, would it be a reasonable assumption to say that there's an economic value of about \$100 billion in those seven sectors in Canada?

• (0905)

**Mr. Michael Chui:** There are a couple of things.

The first question was how we picked those sectors. Those sectors weren't picked because we thought those were the sectors where the most value would be created, but we wanted a number of sectors that varied across a number of dimensions. You notice some of them are B to C sectors, consumer focused. Some of them are B to B, and they're more business focused. Some of them are products. Some of them are services. Some of them are more public services, such as education, and some of them are very commercial.

Really what we wanted to do was to have a variety, and that's how we chose them really. It was really meant to give us a flavour for how open data could work in a number of different types of sectors.

Clearly, there are lots of sectors we weren't able to do as part of the research, and as you said, that suggests there will be even more value potential there.

In terms of trying to size the approximate potential for Canada, it's probably not unreasonable to use the sort of metric that if Canada's economy is this much smaller than the U.S. economy, then potentially Canada would be in that level of magnitude. Of course I wouldn't put any precision around it, but I think that's reasonable.

The other thing to keep in mind is that these are not GDP statistics, because they include consumer surplus, which is not captured in GDP. It's important to make sure, if you're trying to compare, that you wouldn't say this has this much GDP impact, because, as I said, over half of that impact would be a measure that is not captured by GDP. I think that's a flaw in the GDP statistic as it turns out.

**Mr. Bernard Trottier:** Thanks for that clarification.

Some witnesses we've had before this committee talked about the sources of value. There's something of a focus on development of new applications, but others have said that really the true source of value, when it comes to open data, is the sense of removing friction in interactions between different stakeholders in the economy.

You could go with the example of the old days, even within the same institution or within the same government, for example. In the old days of making a request for data, you had to wait several days for that data request to be processed, then receive the data, and then translate that data. There's lots of inefficiency built into the older models, and with open data, things are able to move that much more quickly.

Is that a way to capture the primary source of the value in the McKinsey report that was done last fall?

**Mr. Michael Chui:** Yes. As we took a look at it, in fact, we found many different sources of value. Some of them are creating new products and services. Some of them are creating more efficient markets or more efficient ways to get to information.

One of our interesting findings was that while a lot of open data government efforts allow third parties to have access to government data, sometimes they actually allow other government agencies access to government data, which was more challenging.

I would say, as I said before, that about a third of the impact we found comes from the ability to compare, to do benchmarking, to compare your performance as either an individual or a company or a government against that of others, whether in procurement, in operations, or otherwise. There's a lot of benefit from that, and that really comes from just the ability to understand what best practices are in a different place.

**Mr. Bernard Trottier:** What about you, Mr. Baker? In some of the work you do, what would you say are the main sources of economic value that are generated by open data?

**Mr. Paul Baker:** Well, to get back to the previous point, I was recently at a meeting in the City of Chicago with people who work on the open data portal there. I'll give you an example. The department of housing has data.... I mean, there's housing data in seven different departments. Often the right hand doesn't know what the left hand is doing. One of the biggest consumers of Chicago data is actually other departments within the City of Chicago, which is important.

Washington, D.C. was the first city to release substantial data from almost every department, in 2006. That was the first major effort. They studied the actual users of the data. Between 60% and 70% of the people who came to the website and downloaded the data were actually members of city departments. A lot of city departments were afraid to release data initially, but then they ended up finding out it was so much more efficient to be able to get data that they didn't know was related to what they wanted to do from other departments. That's definitely one of the most immediate benefits.

People within government who are kind of reluctant to release data and make their data available, once they see that it has a lot of benefits for them also, it reduces the resistance tremendously. So within government, it has a very immediate value.

In terms of efficiency, I'll give you one example. We were working with the housing department. They have section 8 housing vouchers where basically low-income people can get money to rent apartments, and landlords or people who build apartments can get deferred taxes for eight to ten years in return for renting to low-income people.

It was a paper-based system, so when someone would leave section 8 housing, it might take six months for a landlord to rent the apartment. That's not a very good incentive, if you know you're frequently going to have vacancies and it takes a long time to rent the apartment. As a result, you would have people looking for section 8 housing. It would take a long time for them to find it. At the same time, you would have landlords who would have vacant section 8 housing and couldn't find people to occupy that housing, largely because the different departments dealing with housing data and section 8 housing didn't talk to each other much. It was paper-based.

Now that they have it all machine readable, they can reduce that time to two weeks or a month. It really improves the situation for both the landlord and the person or family who wants housing.

• (0910)

[Translation]

**The Chair:** Thank you. I have to stop you here to turn the floor over to Mrs. Day for five minutes.

**Mrs. Anne-Marie Day (Charlesbourg—Haute-Saint-Charles, NDP):** Thank you, Mr. Chair.

I want to thank our guests for joining us.

The McKinsey Global Institute's mission is to help leaders in the commercial, public and social sectors make decisions on important management and policy issues.

Mr. Chui, earlier, you talked about the loss of confidentiality and privacy. We know that the public uses the Internet to obtain information. For instance, people may want to obtain all the information about breast cancer, find out whether land is available or for sale, or obtain information about a specific home. They love getting that type of information.

However, they are not as pleased when that data is passed on to random companies. We had a case recently where the personal information of nearly one million individuals was disclosed. A few companies were involved in that.

Has your organization developed any measures to protect citizens in such situations?

[English]

**Mr. Michael Chui:** There are a couple of things. One is that our aim, as part of the McKinsey Global Institute, is to inform the policy discussion, but just to be clear, we actually don't provide direct policy recommendations. That being said, I think the topic you bring up around privacy of individual data is an incredibly important one, one that people care deeply about.

I don't know that there's a blanket ability to solve that problem, because many people disagree about exactly what data should be used for what purpose. I do think one interesting direction to think about regarding policy is that it is increasingly difficult to control the creation of data, and sometimes even the distribution of data, although I think thoughtful policy can be used to regulate that. Often what people find most objectionable is particular uses of data. It's not that the data exists, as you described it, but when it's used for particular purposes is when oftentimes people find it objectionable.

I think one interesting policy direction to take is to identify the uses of data that we don't want to have happen, and to regulate or legislate those uses, rather than the data itself.

• (0915)

[Translation]

**Mrs. Anne-Marie Day:** Can you tell us whether the U.S. has that kind of legislation to protect citizens' privacy?

[English]

**Mr. Michael Chui:** I can't say that I have a comprehensive perspective, but for instance, there is legislation around the release of video rental records. That was a very specific example of a use of data, and the legislators decided it was not data that should be released. It's not a terribly wide-ranging example, but it is a specific example of where that has happened.

[Translation]

**Mrs. Anne-Marie Day:** Can you tell us....

[English]

**Mr. Paul Baker:** Could I jump in?

[Translation]

**Mrs. Anne-Marie Day:** Go ahead.

[English]

**Mr. Paul Baker:** There are a couple of specific laws in the U.S. related to privacy. One is HIPAA, which is on medical privacy. It was a major issue before we had the Affordable Care Act, because if insurance companies found out you had a particular disease, they would bar you from having insurance, or employers wouldn't hire you because you would be more expensive, that type of thing. It's less of an issue now.

There are also very strict laws around student data privacy. With elementary and high school students, you really can't reveal anything about them. That doesn't mean that data should not be shared among people who are authorized to use and see the data. For instance, there's one initiative we're working on now. In school, one of the best indicators that there's a problem is, for example, if a second-grader doesn't show up at school, and they're not sick, chances are, especially in low-income communities, the family has a problem of one sort or another. What will happen is the child won't show up. Someone from a social service agency will visit him. They'll maybe interview the family for 15 or 20 minutes, and make a decision whether the child should stay in the family or not, that type of thing. They have almost no data about the kid. They don't have any attendance data, grades data. They don't know whether the mother is in a drug- or alcohol-treatment program. They don't know whether the father is there, or if the father is there, whether he has post-traumatic stress syndrome from being in Iraq or Afghanistan, or something like that.

Compare that with a situation where you're driving your car with a broken tail light and you get pulled over by a motorcycle policeman. The policeman takes your wallet, walks back to his motorcycle, and checks you against 64 different databases. Why can't we serve children in the same way and share data privately among people dealing with low-income kids to try to improve their lives?

There are situations where you obviously don't want to share that data with the public, but you do want to share it with agencies, with teachers, with principals so they can help families develop.

[Translation]

**The Chair:** Thank you very much for your answers.

I will now turn the floor over to Mr. O'Connor for five minutes.

[English]

**Hon. Gordon O'Connor (Carleton—Mississippi Mills, CPC):** Good morning, gentlemen.

One of the things we wonder about is whether the federal government is in line with other countries or not, whether they're producing enough data. We have a statistic here that says 193,000 data sets are on the open data portal of the Government of Canada. How does this compare to other countries or other areas?

I'll ask both of you to comment on that.

**Mr. Paul Baker:** The numbers vary. In the City of Chicago everyone always said there were 800 and some data sets, and I just talked to a guy who is in charge of the data sets yesterday and he said there are something like 550 data sets. The Obama administration originally, when they announced data.gov, said there were 190,000 data sets or something like that. Now they've reduced that to 54,000 data sets.

Even though the federal government in the U.S. has released lots and lots of data sets, and there's been a lot of data released for a very long time—weather data, satellite data, road-related data, geographic data—really, going back 20, 30, or 40 years, there's a study recently saying that even with all the open data efforts in the U.S., less than 10% of federal agencies actually release significant amounts of data. Much of that is because they make a fair amount of money by selling data either to other agencies or within their own agency. It's kind of a net wash. One agency is paying another agency for data, and the other agency is receiving income, which doesn't make much sense when you're within the same organization but it's part of the accounting rules that's creating inefficiency there.

We would like to see lots more data. There is a lot of data all over the world. We're members of the international Open Data Institute. We're the Chicago node. The U.K. has made tremendous strides in releasing data. France is releasing quite a bit more data. We have a fellow in our office right now who's in charge of the open data portals for the Dominican Republic, which didn't have a particularly democratic government for a very long time. Even there, there's an alliance of progressives and conservatives who want to release data for budget accountability, political accountability, so you can have an alliance among both sides of the political spectrum. We've been talking with them and learning a lot about what's going on there.

Things are happening. There's an effort in the Philippines, and even Russia is taking steps towards releasing more data. There are about 17 nodes in the Open Data Institute worldwide now. We were the first node just last October, and between then and now there are 16 more nodes. Many of them are in countries that were not democratic 10, 15, or 20 years ago.

● (0920)

**Mr. Michael Chui:** I would just add a couple of comments.

It is actually not straightforward to compare countries in terms of their "number of data sets", because you could easily combine two data sets and say it was one. That's probably not a great metric. It is actually difficult to come up with a good metric for it. Certainly, you don't just want to count the amount of data that's out there. Ideally, you'd want to count the impact that data is having on the economy, which is a much more challenging thing to do. That being said, the Open Knowledge Foundation does have at least one benchmark and they try to compare countries. I wouldn't say I necessarily endorse it, but it's interesting to look at.



Finally, if you'd indulge me, I would go back to the privacy question. I think the approach taken in HIPAA versus at least one aspect of the Affordable Care Act in the U.S. does illustrate the difference between legislating the generation or control of data itself, which is HIPAA, versus the Affordable Care Act where you are not allowed to use data in order to discriminate about who is covered by insurance. That's basically the difference, whether or not you're allowed to have data at all or how you use it. I think that's a difference in the way you can legislate policy with regard to privacy.

[*Translation*]

**The Chair:** Thank you, Mr. O'Connor. Your time is up.

I will now give the floor to Mr. Byrne for five minutes.

[*English*]

**Hon. Gerry Byrne (Humber—St. Barbe—Baie Verte, Lib.):** The conversation seems to be somewhat dominated by economics and economic capacity and opportunity. I think there are two stovepipes in this discussion. One is the economic opportunity and the other is civil society and their expectations that open data could lead to new transparency and new information being involved.

Do either one of you see a conflict in terms of the progression, the evolution, of open data? Is there one stovepipe over the other that seems to be dominating the evolution of open data? Are there pitfalls or concerns that either one of you had? I say this because you both come at this from somewhat different perspectives, not disharmonious, but different perspectives. I'd love to receive your perceptions on that question.

Mr. Chui, would you like to begin?

**Mr. Michael Chui:** Sure.

First of all, I think the use of open data to create transparency and accountability in our public institutions is incredibly important. We focused our report on the economic potential, partly because we thought that was part of the dialogue that hadn't been explored quite enough. In addition, we focused it because of our expertise as part of a management consulting firm doing economics and business research. Certainly we agree. I don't think there's conflict necessarily. I'm not sure I'd describe them as stovepipes, but rather as different benefits that you can obtain from open data. Certainly we believe that accountability and transparency are incredibly important as well. We don't think that deriving economic value reduces the potential impact on civil society.

• (0925)

**Hon. Gerry Byrne:** Thank you.

Mr. Baker.

**Mr. Paul Baker:** The Open Data Institute's Tim Berners-Lee founded the group, and he invented the web. The slogan he came up with was: knowledge for everybody. This idea is it's not just the government leaders, not just the corporate leaders who should know intimately what's going on, but ordinary people should have access to similar levels of information to make democratic decisions, to be informed when they vote or participate in politics. From ODI's point of view, it's both the democratizing aspects and the economic benefit.

The Open Data Institute aggressively supports businesses using open data to create businesses partly for social good. There are some major problems, things like climate change. We're working on a project now to link Arctic researchers so they can share information more quickly, and we can make more progress on researching climate change, sea ice floes. Sea ice is melting very quickly, glaciers are, but if we can accelerate the research process by sharing data, space on the icebreakers.... It costs \$50,000 a day to rent an icebreaker. If you can have three or four different teams renting space for a few days at the same time, you can cut down the cost, that type of thing.

There are many areas. Obviously, we've taken a lot of steps backward with the Billionaires United decision in the U.S., where a lot of secret money is now going into politics, and it's a serious problem, but there are many public issues. For instance, there's a lot of controversy around charter schools in the U.S. A lot of groups, like the League of Women Voters, the PTA, are trying to discover data to try to figure out whether charter schools are doing good or bad. The way a lot of the laws are written, charter schools don't have to release as much data as the regular public schools do, so it's very hard to compare apples and oranges. On one side, on the charter school side, you have a big PR effort going on, sponsored by millionaires again, who want to take a certain portion of the public education budget. On the other side, you have teachers and parents who like the public schools and want to keep them. We don't have knowledge for everybody here. The charter school operators know what they're doing. They don't reveal it. We're also involved with Common Cause in another effort to try to uncover some of that data.

I would say both are very important. It's hard to know which is more important.

**The Chair:** Thank you, Mr. Byrne.

[*Translation*]

Your time is up.

We now go to Ms. Ablonczy for five minutes.

[*English*]

**Hon. Diane Ablonczy (Calgary—Nose Hill, CPC):** Thank you, gentlemen, for sharing your expertise with us.

As you can tell, we don't just want to put out a lot of information. That's part of it, but we're struggling with how to focus that into positive results, to value-add for government, for business, for organizations. That's why we're particularly interested in your studies.

Can you boil it down for us? If you wanted us to focus on the best results you think could be garnered from the open data project, and how we could foster them and how we could incentivize people to use them, what would you say are the biggest bangs for the buck?

● (0930)

**Mr. Michael Chui:** As an advocate of using data, I think it would potentially be a mistake for me to throw out ideas without actually doing any analysis. I think the thing you would want to do is to take the type of analysis that we did here globally and apply it to Canada and try to understand what types of data could potentially create the most value in Canada.

As we looked at it we did find, in transportation for instance, tremendous potential. In education we found tremendous potential. What you would need to do is to try to understand, from a Canadian perspective, where the most bang for the buck would be. I think it would be a mistake for me to speculate without having actually done some analysis, given that we actually believe in data as the way to make those decisions.

That being said, I do think the other insight you had was incredibly important, which is that “field of dreams” doesn't work by itself. You make the data available and benefits occur. You actually need to create a vibrant ecosystem of users of data. What that means is engaging with people who will develop programs, develop apps that actually use the data that make it useful to companies and individual citizens, etc.

As I said before, that's almost a marketing game. What we've seen here in the United States, the government has done things like create events. I know that in Canada, for instance, I was in Toronto for the CODE hackathon. It was one of those events that made it more noticeable to people who write computer programs that in fact there's a vast source of data they can use to create new applications. Events and contests, even advertising, quite frankly, are some of the things that have to happen in order to create an ecosystem of, let's call it loyalty. The customers of the data who are the developers of applications have to first, be aware, and second, incented to create applications using open data.

**Hon. Diane Ablonczy:** Paul, do you have any thoughts on that?

**Mr. Paul Baker:** Yes. I like to look at the really big issues, which in my opinion are climate change and medicine right now, at least two of the kind of scientifically oriented areas. There's tremendous activity there. In the climate change area the U.S. federal government has mandated that anyone who gets federal money for scientific research has to release their data within a year. That's fairly recently. Informally it has been going on for a couple of years now, but people haven't adhered to that policy particularly well. From now on, I think they're going to adhere and that is going to accelerate the pace of scientific research, including on climate change which, at least for people who believe climate change is happening, is extremely important.

We see a tremendous amount of energy also in medicine. Silicon Valley is now “disrupting”—an overused word—medicine in many ways. The federal government and even some drug companies are sharing data about their drug studies. A couple of companies have committed to releasing all of their data related to the drug studies they've done, which could help treatments and could also help better

figure out what medicines actually work and what medicines don't work.

There's an effort in the U.K. where half a million people have agreed to have their genes analyzed so they can combine the genomic data with their electronic medical record data. Kaiser Permanente in the bay area has a million of their patients agreeing to have their genes analyzed and combined with their EMR data. They're opening that data up to qualified researchers who follow privacy procedures. The promise in that area is simply tremendous.

There's also crowd-sourced medical data which is really interesting. For instance, some researchers and doctors in California have created a mobile phone app where, if you have what you think might be a cancerous mole, you can take a picture, and that picture goes into a database and then when the mole is removed and analyzed, you come back and the application tells you whether it was cancerous or not. They've accumulated enough data now that by simply taking a picture of a mole, you can get a pretty good probability as to whether you should have it analyzed or not. They're using artificial intelligence to analyze the colour, the pattern on the outside, the size, all that kind of stuff. They're building up crowd-sourced ways of doing this.

A lot of people have lots of moles, but if you have a mole you're worried about and think you should get it tested, you're much more likely to get it tested. Every time you test, it's another \$1,000.

● (0935)

[*Translation*]

**The Chair:** Thank you. I have to stop you here, Mr. Baker.

It is now Mr. Ravignat's turn.

**Mr. Mathieu Ravignat:** I thought the Chicago Open Data Institute projects were especially interesting. You have carried out many of them. Congratulations on your research and your analyses.

I would like to talk to you specifically about the Chicago Lobbyists project. Lobbying at all levels of government is a concern for me.

In your portal, where does the information you have gathered come from?

[*English*]

**Mr. Paul Baker:** It comes from the City of Chicago. The City of Chicago actually had been collecting lobbying data for quite some time. Under Mayor Daley they just didn't release it, and then when Mayor Emanuel came in, he released it within the first couple of months of his term.

There were initially 14 different data sets, while now there are 17 or 18 different data sets. There was a handful of us who had been advocating for open data for quite a few years. During the Daley administration, we hadn't really been able to get a significant amount, and when Emanuel starting releasing stuff, we were complaining once he actually gave us some data that it was incumbent on us to actually do something with it.

I started looking for data sets we could do something with. I saw these 14 lobbying data sets, but if you look at one data set, no one can learn anything from it, so we basically got volunteers. We got a couple of guys from Groupon together, three people from Webitects, a couple of volunteers. One woman was riding her bicycle across the country, going to work for Code for America in San Francisco. She stopped in the city and Google sponsored a hackathon. We got this team together. We went from seven o'clock in the morning on the hackathon, and by seven o'clock in the evening, we pretty much had it done. Some were really high-quality designers and developers, and we all worked together, but the interesting part—

[*Translation*]

**Mr. Mathieu Ravignat:** I have to interrupt you to ask another question.

Has the government shown more transparency in that area? Has data been used effectively to protect the public and reduce the number of conflicts of interest?

[*English*]

**Mr. Paul Baker:** Well, we've spent a fair amount of time looking through the data, and the vast majority of the data doesn't implicate anyone. Much of the data is about a non-profit hiring a lobbyist to help them get a building permit to add an addition onto their building, or businesses that want to get building permits, or someone who has done a study and they want to influence spending by the city. It's that type of thing. There's very little that indicated any kind of corruption.

One of the main problems we wanted to answer was, Walmart had been trying to get a Walmart in Chicago. They were basically banned from Chicago, and so they found an alderman, or someone who was running to become an alderman. They thought that if they could get him to support Walmart.... They gave him a lot of campaign money. Between them and the lobbyists, and other people, he got something like \$300,000 to run his aldermanic campaign, and his opponent only had about \$30,000. We traced all the lobbyists that Walmart hired, every committee they appeared in front of, but we didn't have the campaign contribution data then. We now have the campaign contribution data. We'd like to combine that and try to start looking for political influence.

Walmart was successful. They got the guy elected, and they got a Walmart. Now they have several Walmarts throughout the city.

[*Translation*]

**Mr. Mathieu Ravignat:** Is the fact that no embarrassing situations involving the government have been uncovered in terms of data lobbying due to the nature of the data? Should there be more data or different data so that abuses can be detected, or is the current data sufficient?

● (0940)

[*English*]

**Mr. Paul Baker:** The ethics commission of the city asked us to testify in front of them on how the lobbying data could be improved.

One of the big problems with it is that it's only released annually, so by the time you actually get the data, most of the decisions have already been made. We recommended that it be released in a much more timely fashion: daily, weekly, or at the latest, monthly. That would make big improvements.

Also, we're only told which committees they appear in front of, or which people they talk to; we're not told on what. There's not enough detail on what issue they're talking about. The lobbyists are basically allowed to describe what they're doing with almost no details.

We've also talked with the City of San Francisco about the same issue. We're even thinking about putting together basically a national software application to collect and show the data, one that has more granularity, faster reporting, and all that kind of thing.

[*Translation*]

**Mr. Mathieu Ravignat:** That's very interesting. Thank you.

[*English*]

**The Chair:** Thank you for your answers.

Mr. Aspin, you have five minutes.

**Mr. Jay Aspin (Nipissing—Timiskaming, CPC):** Welcome, gentlemen. Thank you for helping us with our study.

Mr. Baker, you've made reference to the international Open Data Institute. Just recently, I think, our own Canadian Open Data Institute began. The information I have here is that it began as of October 2013, so it's virtually at its starting point.

Sir, what advice would you give our Open Data Institute in terms of best practices on a go-forward basis, given the fact that you, I think, have been at this for at least seven or eight years?

**Mr. Paul Baker:** Actually, we talked directly at a meeting in London just a few months ago, and we talked on the phone maybe two months ago.

One of the purposes of the Open Data Institute is to create this network where everyone can learn from each other. I don't know a heck of a lot about Canada, but I know more than I did before. They are just getting started. Also, at an Open Data Institute event about a year ago, there was a representative from the government whom I talked with for quite a while. I don't remember his name now, but he was in charge of thinking about open data policy. We had conversations there.

In terms of advice for ODI Canada, it has taken a long time in Chicago and in the U.S. to really get the open data movement going. It has been driven by enlightened government leaders, but also by activists, designers, developers, and political transparency people. It's a multipronged process that takes quite a while. I'm glad you're having these hearings, because these hearings.... It's very hard to have an open data movement without any open data, so you have to get some open data from states, cities, and the federal government, and then it has to be non-trivial. It has to be stuff that you can really work with.

Up there, let's say in Toronto, I'm guessing there's probably transportation data available, whereby people can create bus trackers and train trackers and that type of thing, to try to better schedule their commute to and from work. If there isn't, then.... You have to figure out what is the low-hanging fruit, what you start with, and how you work with government and private developers and designers to get things together and get things going. You need a moderately sized town to really have enough data, enough political support, and enough developers and designers to work on it.

You have to work slowly and target your efforts. Canada is a big place with a relatively small number of people. My mother was actually born in Canada and I've spent a fair amount of time there. Now you have a better-off middle class than we do, so maybe you'll have a lot of demand for the benefits that open data can bring.

• (0945)

**Mr. Jay Aspin:** Mr. Baker, there's another project of the Chicago Open Data Institute that I think you've referred to as Chicago lobbyists, and it aims to increase government transparency. I wonder if you could briefly tell us what kind of information is made available by this particular project.

**Mr. Paul Baker:** Do you mean that particular project?

**Mr. Jay Aspin:** Yes.

**Mr. Paul Baker:** It basically lets you look at all the lobbyists, who receives the greatest amount of money, which companies or organizations they work for, what their goals were, generally speaking, and who they talked to within the city.

The question we always posed was this. Who did Walmart hire in order to make it easier to get a Walmart in Chicago? Who did it talk to? Who were the people it thought were important politically within the city to talk to?

Now we didn't have the political donation data in until recently, so we don't know how much money was actually spent, but we at least know.... It's not just Walmart; we want to be able to track any kind of organization, and we can track what conversations they've had, within which departments.

Often, lobbyists are hired to make presentations about a particular topic. We don't know exactly what their presentations are, but if they're suspicious, you can go back and look through the open records data and find details. We don't have that open records data integrated in there either.

**Mr. Jay Aspin:** Thank you, Chair.

[*Translation*]

**The Chair:** Thank you, Mr. Aspin.

Our hour with this morning's two witnesses is up. I want to thank them for their time and for making themselves available to us. I am sure this will help the committee members in the remainder of their study on open data practices in a Canadian context.

I will now suspend the meeting for a few minutes, after which the committee members will reconvene to discuss committee business.

Thank you both and have a good day, in Chicago and in Miami.

[*The proceedings continue in camera*]

---







Published under the authority of the Speaker of  
the House of Commons

---

### SPEAKER'S PERMISSION

---

Reproduction of the proceedings of the House of Commons and its Committees, in whole or in part and in any medium, is hereby permitted provided that the reproduction is accurate and is not presented as official. This permission does not extend to reproduction, distribution or use for commercial purpose of financial gain. Reproduction or use outside this permission or without authorization may be treated as copyright infringement in accordance with the *Copyright Act*. Authorization may be obtained on written application to the Office of the Speaker of the House of Commons.

Reproduction in accordance with this permission does not constitute publication under the authority of the House of Commons. The absolute privilege that applies to the proceedings of the House of Commons does not extend to these permitted reproductions. Where a reproduction includes briefs to a Committee of the House of Commons, authorization for reproduction may be required from the authors in accordance with the *Copyright Act*.

Nothing in this permission abrogates or derogates from the privileges, powers, immunities and rights of the House of Commons and its Committees. For greater certainty, this permission does not affect the prohibition against impeaching or questioning the proceedings of the House of Commons in courts or otherwise. The House of Commons retains the right and privilege to find users in contempt of Parliament if a reproduction or use is not in accordance with this permission.

---

Also available on the Parliament of Canada Web Site at the following address: <http://www.parl.gc.ca>

Publié en conformité de l'autorité  
du Président de la Chambre des communes

---

### PERMISSION DU PRÉSIDENT

---

Il est permis de reproduire les délibérations de la Chambre et de ses comités, en tout ou en partie, sur n'importe quel support, pourvu que la reproduction soit exacte et qu'elle ne soit pas présentée comme version officielle. Il n'est toutefois pas permis de reproduire, de distribuer ou d'utiliser les délibérations à des fins commerciales visant la réalisation d'un profit financier. Toute reproduction ou utilisation non permise ou non formellement autorisée peut être considérée comme une violation du droit d'auteur aux termes de la *Loi sur le droit d'auteur*. Une autorisation formelle peut être obtenue sur présentation d'une demande écrite au Bureau du Président de la Chambre.

La reproduction conforme à la présente permission ne constitue pas une publication sous l'autorité de la Chambre. Le privilège absolu qui s'applique aux délibérations de la Chambre ne s'étend pas aux reproductions permises. Lorsqu'une reproduction comprend des mémoires présentés à un comité de la Chambre, il peut être nécessaire d'obtenir de leurs auteurs l'autorisation de les reproduire, conformément à la *Loi sur le droit d'auteur*.

La présente permission ne porte pas atteinte aux privilèges, pouvoirs, immunités et droits de la Chambre et de ses comités. Il est entendu que cette permission ne touche pas l'interdiction de contester ou de mettre en cause les délibérations de la Chambre devant les tribunaux ou autrement. La Chambre conserve le droit et le privilège de déclarer l'utilisateur coupable d'outrage au Parlement lorsque la reproduction ou l'utilisation n'est pas conforme à la présente permission.

---

Aussi disponible sur le site Web du Parlement du Canada à l'adresse suivante : <http://www.parl.gc.ca>