Introduction
Never before has there been more data, nor more advanced analytical techniques with which to analyze and interpret information. Data holds enormous potential for discovery and innovation, and promises to help better the lives of Canadians, improve the delivery and results of government services, and grow the economy. There are vast opportunities to use data for good, particularly if we develop the long-term infrastructure so that data are preserved, interoperable, and can be accessed responsibly.
On February 23, 2017, four government departments — Global Affairs Canada, Indigenous & Northern Affairs Canada, Public Health Agency of Canada, and Innovation, Science and Economic Development Canada — partnered with Facebook to workshop the most exciting opportunities to use data, analytics, and technology to transform government planning and service delivery. The four partner government departments, as well as representatives from the Privacy Commissioner’s Office, the Privy Council Office, the Treasury Board Secretariat, Statistics Canada, the Department of Canadian Heritage, the Assembly of First Nations, the First Nations Information Governance Centre, the First Nations Health Managers Association, Carleton University, the GovLab at New York University, and UNICEF, participated in workshops with Facebook data scientists to discuss ways to turn policy challenges into data-driven solutions.

In the private sector, organizations have identified ways to strengthen their businesses using online data. The goal of this workshop was to apply these learnings to problems that exist in the public sector. The workshop participants collaborated with Facebook data scientists to assess whether data-driven solutions could help achieve diverse government goals, ranging from understanding the demographics of those accessing consular services in danger zones, encouraging STEM post-secondary education, measuring digital literacy in the Canadian workforce, and taking steps towards reconciliation with Indigenous peoples. The problem statements touched a wide variety of issues, segments of the population, stakeholders, and potential data sources. Through the series of workshops, the participants took steps towards creating concrete action plans to help each department leverage data to more efficiently and effectively address their policy challenges.

"Each government department was dealing with a similar research challenge: identifying the right people to receive a message, reaching those people with the right message, and then assessing whether that message led to a desired effect.”

Key Principles for Data and Policy Development

Participants developed six principles to inform the responsible use of data for good:

1. **MOST POLICY CHALLENGES CAN BE ADDRESSED THROUGH A COMMON RESEARCH PARADIGM**

One of the major learnings from the event was that each participating government department, despite their diverse missions reflected in the discussion, was dealing with a similar research challenge: identifying the right people to receive a message, reaching those people with the right message, and then assessing whether that message led to a desired effect. Of course — identifying the right audience and targeting them with a message that resonates is not an easy challenge — but it is one that becomes tractable with a clearly defined research question, good data and a sound research design.

2. **PARTNERSHIP IS KEY**

More partners means more expertise and, likely, better data. There was a shared understanding that early identification of the right partners and their engagement at all stages of the research was crucial to ensure project success. But, a common question was: How do you get started?

The first step is to create a Stakeholder Map: Focus not just on those partners who are interested in the outcome of the research, but also those who have the data to uncover new areas of intersection and innovation. Potential partners should be evaluated based on the data they can provide, their likely contribution to solving the research question, and the rigor of their data collection and privacy practices. It was also noted that often the right partners are just as likely to be outside of the government as they are to be within the government.

When proposing a collaboration, think about what each partner can bring to the table, and find shared incentives. Furthermore, when approaching a partner, it is important to clearly articulate both the importance of your research question, and how contributing will result in a mutually beneficial outcome.
3. REPURPOSE & REUSE

Think creatively: the data you need may already exist but have been collected by a different organization for an entirely different purpose. Consider stakeholders who may have different incentives to collect the data you need, and how you could work together (see Reaching the Right Audiences).

4. PRAGMATIC PRIVACY FOR RESPONSIBLE INNOVATION

In addition to using data to promote progress and innovation, the onus is on organizations to use data responsibly and in a manner that preserves privacy. That means developing clear guidelines on how to evaluate the benefits of research against potential downsides.

Many participants commented that there was a need for a pragmatic framework to guide practical data analysis. Privacy standards cannot exist in a vacuum, developed separately from the potential applications of data and the benefits that they can bring. Rather, guidelines should be actionable, steering the collection, storage, and analysis of data, as well as the dissemination of insights (see A Call for More Pragmatic Privacy Frameworks).

Participants aligned on four principles for responsible research:

1. **Benefits & motivations**: Ensure that the research provides real benefits to people, community, and society.
2. **Potential downsides & mitigations**: Identify potential downsides, and mitigate them.
3. **User expectations**: Conduct research that is consistent with people’s expectations for how their data is being used.
4. **Data privacy & security**: Implement controls to protect and secure data.

**A Call for Pragmatic Privacy Frameworks**

The need for pragmatic privacy frameworks was echoed by keynote speaker, Matthew Mendelsohn, Deputy Secretary to the Cabinet, Results and Delivery, in the Privy Council Office:

“One of the reasons Canada can be a leader is because we believe in openness and have privacy protections. That discussion is a real one and we have to have balances there. I do think that we are at a time, as we start to dig into data holdings, that we can look at how we ensure that our privacy legislation is modern and appropriate and facilitates data sharing where appropriate. We do not want to be in a position where we have fallen too far on one side of that and prevent the ability to do the work that needs doing.”

**Reaching the Right Audiences: Global Affairs Canada**

Global Affairs Canada is trying to understand the number of Canadians that travel to conflict zones. Statistics Canada collects information (through customs forms) about where Canadians have traveled. While Statistics Canada’s goal is not to measure the percentage of Canadians traveling to conflict zones, the same data they collect could be reused for this purpose and thus serve the needs of Global Affairs. Since each department’s reasons for collecting travel data are different, Global Affairs had not previously considered a partnership with Statistics Canada. Systematically evaluating their research question and data needs pointed to Statistics Canada as a potential partner with promising data.

Workshop participants discuss key principles for using data to tackle policy challenges.
5. BE CREATIVE AND FLEXIBLE WHEN DEFINING DATA NEEDS

At the beginning of any research project, it can be tempting to ask for the most comprehensive and granular dataset available. If these data are easily obtained — great! Innovation is largely driven by the discovery of new purposes for previously collected data.

However, collecting the “ideal” dataset can sometimes require tradeoffs in time, expense, and resource capacity (for storage and analysis), as well as have implications for privacy. For these reasons, it is important to be precise and flexible when defining data needs.

A few tips can help:

• **Define your unit of analysis:** Many policy problems will not be addressed at the individual level or in real time. Think about the level of aggregation at which decisions will be made (across time and space), and whether the data you need exist at that level.

• **Focus on layering data rather than finding the data holy grail:** Rarely will you discover the “perfect” dataset for your problem. Rather than getting stuck, think about how you can layer data from multiple sources to create a more complete picture.

• **Consider alternative data flows:** Collecting and analyzing data can help you understand a problem and inform your strategy for addressing it. However, you do not always need to possess data in order for those data to be put to work (see The Value of Partnership).

---

**The Value of Partnership**

In working to reach Canadians traveling to conflict zones, many participants thought it would be helpful to send direct messages to people in certain regions. Initially, their instinct was to ask for geolocation data from an Internet or telecommunication provider for each individual traveling to a conflict zone, thus focusing on data for individual people. But, the sharing of sensitive geolocation data raises important privacy concerns. Fortunately, the participants also concluded that they did not need this information in order to achieve their objective of reaching Canadians in conflict zones. Participants found that they could instead partner with companies to use the same data but in an aggregated and anonymized way to send public service messages to people in dangerous environments without the need to work with individual data points. There is precedent for this approach: for example, in partnership with governmental agencies, phone companies send texts to people in disaster zones. Additionally, Facebook delivers AMBER Alerts to people’s News Feeds in targeted search areas after a child has been abducted. In this way, the partner is able to achieve its policy objective without actually needing access to personal information.

---

The workshop was led by data scientists, like UNICEF’s Emily Garin, who started by helping each government department define their problem.
6. BUILD INFRASTRUCTURE FOR DATA COLLABORATION

Even if the dataset you need exists, it may not be easily accessible. In fact, one of the biggest barriers to data-driven research is internal processes that make it difficult to begin research projects, convene the right partners, and demonstrate the marginal benefit of the research to the organization or to society (see A Chicken & Egg Problem).

Recommendations to overcome this friction include:

- **Make interoperability a priority:** Governments are investing resources in Open Data offices that seek to build interoperable datasets and computational systems across governmental entities. These offices can ensure that data sets are accessible across organizations, and act as an independent resource when assessing the risks and tradeoffs of sharing data across organizations. Of course, any interoperable system should be created in a privacy protective way. One such example is Canada’s Open Government initiative and the Third Biennial Plan to the Open Government Partnership. The main goals of the partnership are to make data open by default and encourage fiscal transparency in order to enable innovation, prosperity, and sustainable development.

- **Build internal literacy:** Policymakers and data scientists/analysts do not always speak the same language. For this reason, it is necessary for organizations to build internal literacy around the process of sharing data and the value of data-driven research. Internal literacy will help establish clear channels of communication, while ensuring that both sides understand the purpose of solving a research problem. For an example of this approach, see Building Cross-Functional Capacity: 18F.

A Chicken & Egg Problem

Participants discussed the value of the Database of Children with Sport Injuries — which is currently only accessible by Health Canada. The conversation focused on the difficulty and cost in terms of resources and time to open access to the database for use across all government institutions. For instance, in order to open the dataset, it is necessary to demonstrate the dataset’s impact; however, to demonstrate the impact, you need to have access to the data — a classic chicken and egg problem. This unclear path to accessing and proving the value of a dataset disenfranchises researchers who do not have the time or resources to navigate the process. A governmental organization focused on building internal infrastructure for data collaboration and helping policy experts articulate the value that data can bring to their work can help with both interoperability and internal literacy.

Building Cross-Functional Capacity: 18F

A great example of this approach is the work of 18F, an office within the United States General Service administration. 18F works cross-functionally with federal agencies to deliver digital services and tools. For example, 18F aids agencies in understanding and implementing new laws or mandates, and digitize and streamline internal systems to reduce risk and save time. Consistent with this principle, in Canada, Budget 2017 acknowledged the importance of creating more internal capacity for digital services within the federal government.
CONCLUSION

The majority of participants agreed that the key to unlocking their policy challenge was (1) identifying the right audience and targeting them with a message that resonates. In order to do so, it is crucial to (2) identify the right partners to ensure project success. The workshop participants greatly benefited from collaborating with other government departments and external organizations. It is also helpful to (3) be precise and flexible when defining data needs and soliciting partners to contribute data to solve the issue. It is important to keep in mind that the onus is on organizations to (4) use data responsibly and in a manner that preserves privacy. Another helpful finding was to (5) think creatively, as valuable data may already exist that was collected by a different organization for an entirely different purpose. Finally, it is crucial to build (6) the internal infrastructure to support data collaboration within departments, as well as with external organizations.

The six valuable learnings outlined in this report were drawn from an afternoon of workshop discussions. This is the beginning of an important conversation about the use of data to solve the challenging policy issues faced by governments and society today. We are open sourcing the workshop workbooks (see Appendix A), with the goal that they will help other departments translate their policy problems into data-driven solutions, continue the discussion, and learn from each other’s experiences. We hope this process will be helpful to the Public Service of Canada as a whole, as it moves towards leveraging data to address policy challenges.

We are grateful to the partner government departments, Global Affairs Canada, Indigenous & Northern Affairs Canada, Public Health Agency of Canada, and Innovation, Science and Economic Development Canada, for leading the way and demonstrating the potential of data to solve the challenging questions with which they are seized. We are also grateful to the attendees from all participating governmental and non-governmental organizations for sharing their expertise, experience, and insights throughout the workshop. Everyone’s willingness to share and their keen participation were instrumental to the event’s success and the learnings contained in this report.
Appendix A

Refining the Research Question
WORKSHEET

The purpose of this worksheet is to (i) refine the high level problem statement into a well-structured research question and (ii) identify potential data sources that could be used.

Choose your research adventure:

1. State your problem in one sentence.

2. Why is this important to study?

3. Name the main stakeholders whose lives will be made better as a result of this research.

4. How will you measure the success of the research? Define three objective metrics for success.
Refining the Research Question
WORKSHEET CONTINUED

5. Is the goal of your research question to (A) measure a single variable or (B) to look at the relationship between variables?

6. At what level of aggregation do you want to measure these variables across time and space?

7. List all the variables you need to measure.
Refining the Research Question
WORKSHEET CONTINUED

The purpose of this worksheet is to (i) refine the high level problem statement into a well-structured research question and (ii) identify potential data sources that could be used.

8. For each variable, fill in the following information:

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Variable 1:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Variable 2:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Variable 3:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Variable 4:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Variable 5:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Rubric for Responsible Research

WORKSHEET

1. Benefits & Motivations: Ensure that the research provides real benefits to people, community, and society.
   a) Who will benefit from this research?
   b) If you had access to the data you would like, how would this be used to provide a better service to your beneficiaries?
   c) How can these benefits be communicated to the beneficiaries?

2. Potential Downsides & Mitigations: Identify potential downsides, and mitigate them.
   a) Does the research change people’s experience, and if so how?
   b) Will the collection of these data have downsides for people?
Rubric for Responsible Research
WORKSHEET CONTINUED

c) Does your proposed research include vulnerable populations?

d) Describe an approach to mitigating potential downsides and protecting vulnerable populations.

3. User Expectations: Conduct research that is consistent with people’s expectations for how their data is being used.

a) Imagine people found that your org was using data for this purpose, would they be surprised? Do you think you could explain why you are doing this in a way that would mitigate concern?

b) Do you think you could explain why you are doing this in a way that would mitigate concern?
Rubric for Responsible Research
WORKSHEET CONTINUED

   a) Does the data contain personally identifiable information (PII)?

   b) If not, could individuals reasonably be reidentified from the data?

   c) What could be done to make the data less sensitive while still retaining privacy?

   d) Where is the data that you need currently stored?
Rubric for Responsible Research

WORKSHEET CONTINUED

e) If you are not currently in possession of this data, how would you gain access?

f) Will data be transferred between servers? If so how?

g) What are relevant laws and legislation and does your proposal comply?

h) Does your organisation have the internal expertise to responsibly use this data in light of the above? If not, who are suitable vetted third parties?