OpenNorth

COVID-19: Challenges facing open and shared COVID-19 data

MEGAN WYLIE, RESEARCH ANALYST, APPLIED RESEARCH LAB, OPEN NORTH
NABEEL AHMED, PROGRAM OFFICER, ADVISORY SERVICE, OPEN NORTH
JEAN-NOÉ LANDRY, EXECUTIVE DIRECTOR, OPEN NORTH
SARAH GAGNON-TURCOTTE, DIRECTOR, APPLIED RESEARCH LAB, OPEN NORTH

AUGUST 11, 2020

Data has been a key focus of attention for governments and the public alike during the COVID-19 pandemic. Better data has been acknowledged as a crucial factor to mitigate disease spread through initiatives such as contact tracing apps or the concept of 'flattening the curve'. Just as our lives have been impacted in unprecedented ways, so has data's popularity sky-rocketed unexpectedly during the pandemic.

The problem is that COVID-19 data is not perfect. By now, there are well-documented issues with the collection and reporting of the data, such as delays in collecting and sharing, inconsistent reporting between places as well as changes in methods. Canadians also can't access the data they need to understand the spread and impact of COVID-19. Calls for information on the demographics of cases in Canada, specifically race or ethnicity variables, still haven't been fully resolved and people continue to express a desire for more accessible data on COVID-19 for decision-making.

> "Each number is connected to a human story; a neighbor, a family business... the numbers allow us to understand a range of situations and allow policymakers and all Canadians to address them."
>
> - **Mr. Anil Arora**, Keynote speaker, and Canada's Chief Statistician

As Canada's leading non-profit specialized in open data since 2011, Open North knows these issues well. In a recent policy brief, we summarized the needs of data practitioners during the pandemic and also collected use cases illustrating the way governments are stepping up and releasing new data to help with response and recovery during the pandemic. This moment, when governments around the world are responding to COVID-19 with rapid decision-making, is a time to double down on open government principles to maintain the trust of the public and resilience when facing future shocks.

THE THREE MAIN CHALLENGES FACING DATA USERS

To address the needs of data users, Open North and the Standards Council of Canada convened over 100 stakeholders on July 2nd, 2020, to discuss the use of data during the pandemic. Taken by the urgency of the pandemic and the rapid response required, governments hadn't consulted the public regarding information and data transparency, nor about whether the available data met stakeholder expectations. Open North's consultation focused on open and shared data regarding disease spread, community impact, and government response, building upon the International Open Data Charter and the Digital Government and Data Unit of the Organization for Economic Cooperation Development's call to convene stakeholders and learn about local COVID-19 data needs.

Nearly half of the participants worked in the public sector, and over a third identified as belonging to either civil society or academia. These are groups of people that are familiar with open and shared data in Canada, including the challenges that data producers and users face as well as the Canadian COVID-19 data response.

The *Open COVID-19 Data* consultation was hosted online, with opening remarks from the keynote speaker and Chief Statistician of Canada, Mr. Anil Arora, and four additional guest speakers. This was followed by small group discussions that provided opportunities for participants to identify available and desired data sets. Those who registered for the consultation also had the opportunity to fill in a pre-event survey identifying their data needs (receiving 40 responses). The findings reported in this

brief summarize the contributions from participants during the small group discussions as well as the survey responses. For more information on *Open COVID-19 Data: Engaging Canadians on Their Data Needs* consultation, see our [blog post](#).

Challenge #1: Participants want more data opened up, and they want to know when it's available

There is a clear demand for greater data availability and, according to participants, there is still a lot of data that could be opened and shared. This is evidenced by over 200 contributions participants added to our collaborative whiteboard platform during the consultation (see Table 1).

The dataset requests varied substantially. Participants requested more data on remote working conditions, people's perspectives on how we should design our urban spaces, as well as requests for data on testing locations. In our analysis, we found 44 unique dataset requests (list available in Appendix A). We list a sample of the datasets in Table 2 as well as some of the remaining questions that participants want to answer with opened up data.

Table 1: Participants posted over 200 contributions to our collaborative whiteboard

	Disease spread	Community Impact	Government Action	Total
Available	37	20	22	79
Not Available	55	51	32	138
Total	92	71	54	217

*This table counts unique post-it note contributions. This table does not represent unique datasets, as there were many repeat dataset requests by participants. In addition, we noted that many contributions overlapped multiple issue areas, signaling the potential for a few well-chosen data sets that could meet the needs of multiple data users.

The most frequently referenced was demographic data related to all aspects of the COVID-19 disease spread and impact. This data is available in other countries, such as the U.S., and has allowed the public to more effectively evaluate the impact of COVID-19 on certain marginalized communities. Other social and demographic data were also highly sought after, namely people's movement, housing, and employment status.

Table 2: A sample of the wide range of datasets participants requested be opened up

	Disease Spread	Community Impact	Government Action
Sub-topic areas	• Revised case data: false positives and negatives • Testing those who are non-symptomatic • Contact tracing • Vulnerability factors • Inter-person interactions • Mask usage and effectiveness	• Race/ethnicity • Disability status • People's anxiety • Public transit • Housing status (e.g. homelessness) • Type of housing (e.g. social housing) • Area of work (e.g. tourism)	• Recipient of a government aid program(s) (e.g. CERB*) • Essential workers & exposure at work (e.g. gig economy workers) • Remote work • Procurement (e.g. type of business) • Spending for indigenous communities
Remaining questions and/or use cases	• What are the long-term impacts of COVID-19 on people's health? • Are people getting multiple COVID-19 tests? • What is the role of culture in disease spread? • How is transportation a barrier to testing? • How does crowded housing impact disease spread and impact? • Where are people congregating or interacting?	• What is the impact of COVID-19 on marginalized groups? • How does the impact compare between different nursing homes or communities? • What is the impact of children being home? • How has public transit usage changed?	• Who is hiring? Where can people find jobs?

Canada Emergency Response Benefit. Full data request list available in Appendix A.

Challenge #2: Long-standing barriers to data use are exacerbated by the pandemic

Participants emphasized the need for better interoperability, data released at lower spatial resolutions, and higher quality data. Their needs are outlined in Table 3.

Interoperability is the ability to integrate multiple datasets, allowing users to join data. Releasing data representing localized areas, or at *lower spatial resolutions*, helps understand the local situation and is more useful for research. Inconsistent *data quality* also makes it difficult for participants to trust released data and information and use it for research and decision-making.

> "[We] need to agree on common identifiers to allow data linkages across datasets for the same individual."
>
> - **Mr. Mark Leggott**, *Guest Speaker and Executive Director of Research Data Canada*

The pandemic has exposed the urgency with which we need overarching data governance that addresses these challenges and facilitates the use of data across jurisdictions.

Table 3: Long-standing barriers to data use

Interoperability	Quality	Spatial resolution
• Common identifiers and semantics for joining data • Common categorization for variables, such as age brackets, across jurisdictions • Consistent methods and formats across datasets and jurisdictions	• Transparency on data quality changes (e.g. false positives) • Timely data that address pertinent questions • Metadata to assess the quality of the data themselves	• Release data at lower geographies, specifically for rural areas and at the neighborhood level • Nationally integrate geographic units where data are released, for consistent coverage • Maintain the privacy of people the data represent

An important example of this was raised by speaker Bonnie Healy. First Nations are not mentioned in the Canada Health Act, making it difficult to create interoperable data-sharing systems between these Nations and other Canadian health data providers.

> *"When we're looking at data systems, we need to recognize First Nations health systems as active players."*
>
> - **Bonnie Healy,** *Guest Speaker and Health Director at Blackfoot Confederacy*

Challenge #3: It's confusing to access data, especially in Canada

In many cases, stakeholders were confused about what data was available and how to access it. To mitigate this confusion, participants suggested that:

- The data should be accessible from one source, such as Canada's open data portal;
- There should be an ecosystem map of where data is already available;
- Data providers should pre-link data to facilitate its use; and
- Data providers should organize data in a format that makes it easy for people to use in the analysis.

> *"[There is] so much information it's hard to make sense of it"*
>
> *"[Il est] difficile d'identifier rapidement les données déjà accessibles"*
>
> - *Participants*

While recognizing the inherent challenges in coordinating data access across multiple governments, data providers need to clearly communicate data, where to access it and from whom in Canada.

NEXT STEPS

Open North's *Open COVID-19 Data* consultation highlighted three challenges in data access and use, moving from discussion into identifying key action areas. It is crucial that governments provide this data and give communities the opportunity to better understand COVID-19's causes and impacts.

> *"This is a dialogue with Canadians … this conversation has to continue."*
>
> *- **Mr. Anil Arora**, Keynote speaker, and Canada's Chief Statistician*

Open North will continue this conversation through our recently-funded project with the International observatory on the societal impacts of artificial intelligence and digital technology (OBVIA) which will explore the application of data governance mechanisms in health research and practice. We look forward to continuing these discussions with partners in the coming months through our research and capacity-building work. These efforts support greater access to quality, timely, and standardized data that protects public health and improves coordinated government responses through effective shared data governance practices in Canada.

Open North is grateful to the participants who filled out our survey and attended our consultation. You provided valuable insights that are captured here as well as in our Appendix.

ABOUT OPEN NORTH
We work with a wide diversity of innovative and connected public administrations and community stakeholders to build their efficient, ethical, and collaborative use of data and technology to solve complex problems. We sustain global peer-to-peer networks for knowledge sharing that improves smart and open governance practices for citizens across Canada and the world.

Sign up to our newsletter
To get our latest news and resources!

APPENDIX A

Stakeholders also spoke of a variety of other data needs that Open North was unable to capture in our consultation brief. Here we list the raw data that was gathered during our event.

Areas where stakeholders want access to more data

1. *Disease spread*
   - Revised case data: false positives and negatives
   - Testing those who are non- symptomatic
   - Contact tracing
   - Vulnerability factors
   - Inter-person interactions
   - Mask usage and effectiveness
   - Facility crowding and usage
   - Immunization
   - Hospital space demand
   - Patient wait times
   - Pharmacy locations
   - Billing fees
     - Régie de l'assurance maladie du Québec (RAMQ) or other government bodies
     - Facilities
     - Healthcare providers (e.g. doctors)
   - Healthcare facility information
     - Names
     - Services offered
     - Number of beds
     - Address
     - Coordinates

2. *Community impact*
   - Race/ethnicity
   - Disability status
   - People's anxiety
   - Public transit
   - Housing status (e.g. homelessness)
   - Type of housing (e.g. social housing)
   - Area of work (e.g. tourism)
   - Workplace impact
     - Productivity
   - People's priorities
     - What do people value?
     - How do they want society to look after the pandemic?
     - What changes should remain?
     - What should go back to the way it was?
   - People's use of spaces: interior and exterior
   - Personal behaviors, perceptions & risk assessment
     - For returning to work
     - Perception of the effectiveness of masks
   - Community well-being indicators
   - Food relief organizations
   - Substance consumption
   - Access to internet and broadband

3. *Government action*
   - Recipient of a government aid program(s) (e.g. CERB)
   - Essential workers & exposure at work (e.g. gig economy workers)
   - Remote work
   - Procurement (e.g. type of business)
   - Spending for indigenous communities
   - Indigenous Nations
     - Resources available to communities

- - Population data
  - Pre-contact conditions
- Government aid for businesses
- Government procurement
  - Type of businesses
  - Owner characteristics
- Government support for Indigenous communities, by Nation
- Reopening plans and procedures in place
- Public washroom facilities
- Hygiene and maintenance procedures
- Compliance with regulations
- Contact tracing apps
  - Where are they in use?
  - What is their risk assessment?
- Clear information on how to access healthcare in each province
- Beneficial ownership

Geographic scales requested by participants

- Rural
- Municipal
- Census data at the:
  - Dissemination Area level
  - Census Division level
- Lower granularity
- National integrated files

Other stated needs

- Metadata with authority and that is used
- Standardized measures imposed by each level of government
- Contextual information
  - E.g. Historical comparisons at each geographic level

- Analysis information for government decisions
  - E.g. analysis of other country's approaches
  - Second wave plans
  - What data is being collected even if it's not being released (e.g. contact tracing analysis)
- Contact information for data stewards
- Source code for contact tracing
- Design of the system and database
- Community generated data

As these data requests are fairly general, governments will need to re-engage stakeholders in order to understand their particular data variables, geographic scale, and other needs.