

Canadian SARS-CoV-2 Genomics Data Portal

Request For Proposals

2021-02-24

Version: 1.6

Preamble

In April 2020, Genome Canada launched the Canadian COVID-19 Genomics Network (CanCOGeN). The goal of the VirusSeq component of CanCOGeN is to perform genomics surveillance by sequencing the SARS-CoV-2 virus genome in up to 150,000 individuals. The objective of this genomic surveillance is to track the virus spread, monitor for variants of concern (VOCs) that might impact transmissibility or disease severity and assist in outbreak analysis to inform public health and policy interventions. This is the first time that a large-scale genomic epidemiology initiative has been used to guide and inform the public health response to a pandemic in Canada, setting the stage for genomics to serve as a core tool for outbreak tracking in future pandemics. Aside from the genomics surveillance that is being done at a provincial and national level, CanCOGeN also aims to ensure the use of SARS-CoV-2 genomic data for research purposes. The Canadian SARS-CoV-2 Genomics Data Portal will be a critical tool to ensure data access to the research community.

Efficient, sustained, and timely data sharing are essential to advance these objectives. Organizations such as the Global Alliance for Genomics and Health (GA4GH), the World Health Organization and other experts alike have emphasized that data sharing becomes of utmost importance in responding to global health emergencies such as COVID-19.

CanCOGeN and Data Sharing

As part of its mandate, CanCOGeN is committed to rapid, transparent, and equitable data sharing practices that meet Canadian privacy and ethics regulations and standards to guide and inform the public health and policy response to COVID-19 and to enable both national and international research.

Broadly speaking, data sharing of viral genomes and metadata in Canada involves two related but distinct processes:

1. **Data sharing to a limited-access national database.** This database, administered by the National Microbiology Laboratory (NML), informs national public health decisions. Provinces who are conducting their own viral sequencing (Quebec, Ontario, Alberta, British Columbia) share their sequencing data and enhanced metadata with the NML. This database also contains data generated in-house by the NML on behalf of the other provinces. A MOU is in the process of being signed to promote data sharing between the NML and the provincial health laboratories as well as between provincial health

laboratories. As of January 2021, about 40% of sequences generated in Canada has been shared with the NML.

2. **Data sharing to open and/or public international databases of data that does not raise privacy concerns.** Data generated by the NML is shared, along with minimal metadata, via international databases, including the Global Initiative on Sharing All Influenza Data (GISAID). The provincial labs will in some cases send their data to NML to then be shared with GISAID or in other cases they share their data directly with GISAID. This sharing contributes to global genomic surveillance of viral transmission. As of January 2021, about 25% of sequences generated in Canada have been deposited into GISAID. GISAID data is not available through a completely open access process. GISAID maintains terms of use for its repository and approves individual requests for data access. This approach has created some challenges for Canadian researchers who have reported difficulties accessing Canadian data that has been uploaded into GISAID.

As outlined in the Data Sharing Barriers memo (December 2020), the CanCOGeN Ethics and Governance Committee identified three broad categories of barriers associated with data sharing: 1) Legal, Ethical, and Administrative; 2) Scientific and Technical; and, 3) Cultural and Motivational. As part of the Scientific and Technical aspects, one of the barriers is the lack of a National Data Portal that will provide access to scientists, epidemiologist and other researchers that could use this data to advance our knowledge and understanding of the COVID-19 pandemic.

Objective

The purpose of this request for proposals is to adapt, or re-purpose, an existing platform to build a Canadian SARS-CoV-2 Genomics Data Portal that will contain genomic sequences and minimal metadata and will facilitate broad access to this data. The Data Portal will be mirrored to GISAID and will contain a centralized, replicated environment for data storage and analysis in Canada. To protect the data and the confidentiality of people diagnosed with COVID-19, the de-hosted viral genome data and minimal metadata contained in the public Data Portal will not include any identifiable personal data. To the extent possible, data from the Data Portal will be deposited into global public databases, such as GISAID, to contribute to international efforts to track viral evolution. The overall objective of the Data Portal is to enhance the value of sequence data by creating an integrated dataset with minimal metadata and a real-time dashboard summarizing all SARS-CoV-2 genomic data available in Canada for the purpose of advancing research and informing public health at a national and international level.

Key elements

1. **Funding Opportunity.** This funding is not for a research project. The objective is to provide a service to the research community by enabling access to the SARS-CoV-2 genomic data and minimal metadata. The provider of this service will have no special

access to the data, for research or other purposes, beyond improving infrastructure within the scope of this project.

2. **An integrated and pan-Canadian Data Portal.** The Canadian SARS-CoV-2 Genomics Data Portal will contain all SARS-CoV-2 sequences (consensus and raw de-hosted) and minimal metadata available in Canada. The solution must be portable across different cloud providers. Integrating Canadian data in a single database is key to advancing research initiatives and to help epidemiologists, modelers and other research groups have direct access to this valuable data – allowing them to produce timely analyses which will aid decision making for the COVID-19 response.
3. **A single submission process.** Instead of uploading data directly to global public databases, such as GISAID, the NML and the provincial labs, and their healthcare partners, will submit their consensus and raw de-hosted sequences and minimal metadata to the Canadian SARS-CoV-2 Genomics Data Portal. Managers of this portal will develop an automated system to further validate data and to submit all available consensus sequences and minimal metadata directly into the global public databases.
4. **Broad access portal.** The Canadian SARS-CoV-2 Genomic Data Portal will be available to Canadian scientists, public health officials and approved third parties. The application should discuss if registration is required (e.g., institutional email). The Data Portal governance will be the responsibility of an Oversight Committee working directly with CanCOGeN.
5. **Data access, ethics and security.** The CanCOGeN Ethics and Governance Committee will advise on the regulatory and ethical frameworks that are needed to share data held in the Canadian SARS-CoV-2 Genomics Data Portal.

Application and Selection Process

1. **Open Request for Proposals.** A request for proposals will be advertised through all CanCOGeN communication channels including the CanCOGeN website and the different committees and working groups that are part of CanCOGeN's Network. The call will be for a specific deliverable or a service and not a research project. Functionality for bioinformatic processing and exploratory analysis or visualization are not within the scope. However, the proposed solution should be interoperable with existing solutions for bioinformatics and analysis.
2. **Proposal.** The proposal will consist of a short document, maximum 5 pages, describing the solution, the team's expertise, the time frame for implementation and the budget. The plan will be free form text and references will not count as part of the page limit. As part of the implementation plan there should be a description of the technical specifications, data flow and relevant data governance processes (e.g., registered

access). In addition, a signature page (form provided) and short CVs of key team members are required.

3. **Selection Criteria.** CanCOGeN's Data Sharing Committee will work with bioinformaticians and other experts in the field to define the key elements of the Data Portal and develop the criteria for selection of the supplier. Selection criteria include:

Implementation Plan

- a. Capacity to implement the Data Portal in a short timeframe
- b. Availability of a relevant Data Portal that has already been developed or can be re-purposed by the team
- c. Application of international standards to ensure interoperability with other Canadian and international COVID-19 initiatives and deployment on all major cloud vendors
- d. Potential for extensibility to support additional functionality over time
- e. Clarity of data flow that captures interactions with provincial labs, institutions, the NML and GISAID
- f. Capability to ensure a complex logical workflow for receiving and submitting data (e.g., validation, accessioning, user authentication)

Team

- a. Team of technical experts with demonstrated success in related Data Portal development, deployment and maintenance
- b. Team of technical experts and regional champions that initially represent at least three provinces including those involved in viral sequencing data generation
- c. Team is multidisciplinary and includes end-user perspective (e.g., epidemiology, bioethics, alternative disease or non-genomics background)

Management and Governance

- a. Commitment and technical mechanism to follow the data access and data use guidelines being developed by the Data Portal Oversight Committee
- b. Capacity to implement data security protocols, tools and measures to ensure the integrity of the Data Portal and protect from cyber attacks
- c. Quality of the Data Portal sustainability plans
- d. If co-funding is provided does it meaningfully contribute to the CanCOGeN objectives and overall sustainability of the Data Portal

4. **Selection Process.** A selection panel comprised of international experts, members of CanCOGeN's Data Sharing Committee, VirusSeq Implementation Committee, and/or CanCOGeN's Steering Committee will evaluate all proposals and select the one that best meets the stated goals and fulfills the selection criteria.

5. **Funding.** There is \$300,000 available for this funding opportunity. No co-funding is required however it is encouraged as it could augment the ability of the portal to be successful. Private sector organizations are eligible to apply. The term is one year or until March 31, 2022.

Timeline

The overall objective is to move this initiative forward as quickly as possible to ensure a Canadian SARS-CoV-2 Genomics Data Portal is in place in the shortest possible timeframe. The call for proposals will be launched on February 24, 2021 with a deadline for proposal submission to CanCOGeN of March 3, 2021. The review panel will meet on March 8, 2021.

Questions regarding the process and submission of the proposal should be directed toward CanCOGeN staff, specifically, Koko Agborsangaya kagborsangaya@genomecanada.ca.