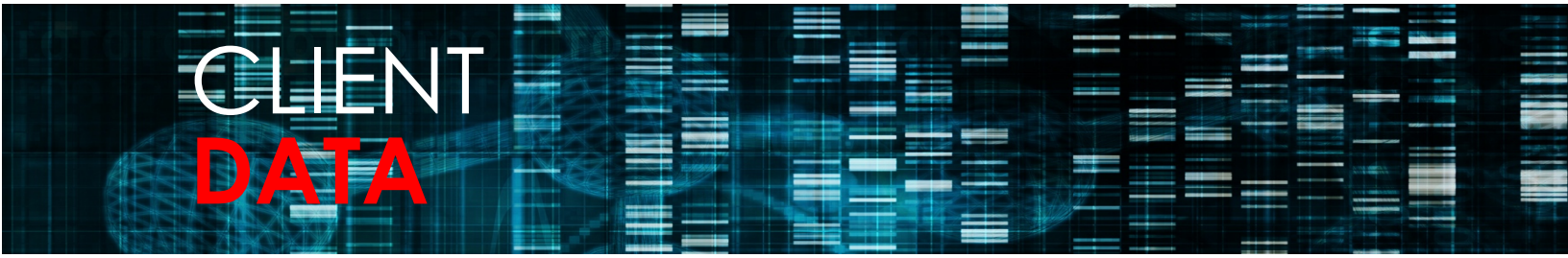
Each CGEn node abides by the information and data security policies, standards and procedures implemented by their host institution's Information Technology (IT) departments and designed to meet federal, as well as their respective provincial, legislative requirements. These version-controlled policies are kept on internal networks, accessible to CGEn staff at each host institution (aspects relevant to CGEn CFI-MSI support are mentioned in the main grant). Topics within the institutional policies include, as examples:

- Information security/cybersecurity
- Network terms and conditions of use
- Acceptable use of assets
- Remote access, mobile devices and teleworking
- Disclosure of personal health information
- Incident and event management procedures
- Risk assessment
- Hardware decommissioning
- Training



Data generated by CGEn on a fee-for-service basis for clients is collected, analyzed and temporarily stored within the framework of the host institutions' data management, cybersecurity and privacy policies.

Physical security is maintained by ensuring computers are kept within locked environments, including data centres containing high performance computing and storage architectures. Portable devices are encrypted to avoid data disclosure in case of loss or theft. Access security is maintained by limiting access to laboratory information systems and network paths to data for analysis to authorized CGEn personnel. Logging and auditing of access capabilities are implemented. Network security is maintained through the use of firewalls, network intrusion monitoring and virus scanning software. To address privacy concerns, CGEn requests the minimal

identifiable data be shared with CGEn staff. Only identifiable data absolutely necessary for the generation or analysis of data is to be collected by CGEn.

Data generated for CGEn clients is considered confidential and is only shared with the scientist (or designate) for whom the service was performed. Once transferred to the client, project data is deleted from CGEn's servers. Method of data transfer is dependent on the data type and file size. Secure ftp servers are most commonly utilized. Transfer of data to a cloud platform requires approval of that cloud by the institution's CISO/CIO following an analysis of the cloud environment's security.

This Governance Framework aims to provide oversight mechanisms for the administration, custodianship and sharing of **data managed** by CGEn. Following the model developed for the COVID-19 HostSeq project, this framework sets out considerations for the deposit and storage of data in a CGEn Databank (e.g., consent requirements, description of privacy and security protections, closure/transfer of the database, management of participant withdrawal, risks, and benefits).

This Framework is intended to assist scientists with potentially eligible collections and their local Research Ethics Boards (REBs) in assessing the oversight mechanisms in place to ensure appropriate data sharing via any CGEn Databank. It also informs researchers accessing the resource about the overall access process.

Mandatory Commitments

As part of contributing data to a CGEn Databank, researchers should note that the following commitments are mandatory (and will be part of institutional agreements):

- A copy of sequence data generated at CGEn will be added to the CGEn Databank (contributing research team will receive their own copy of this data)
- Relevant clinical/phenotype data will need to be provided and will be stored in the CGEn Databank
- Access to the full CGEn Databank (including genomic and clinical data contributed) by approved researchers will be governed through applications to the CGEn databank Data Access Compliance Office (DACO)

Governance and organizational structure

Scientific and Administrative Oversight

The CGEn Executive Committee provides scientific and strategic oversight for the CGEn Databank, in order to optimize scientific outcomes and benefits to Canadians.

Ethical Oversight

The collective CGEn Databank would include data from collections contributed by CGEn affiliated researchers.

For retrospective collections, additional approval will be sought from the local institutional REB to allow data from the collection's participants to be included in the CGEn Databank, according to the conditions established in CGEn's Data Governance Framework (this document). For prospective collections, local REB approval for inclusion of participants' data in the CGEn Databank should be sought at the time of initial ethics review submission to the local REB responsible for ethics oversight of the respective collection, and information about contribution to the CGEn Databank should be provided in the participant consent forms.

Data access and release oversight

Data from the CGEn Databank as described in Table 1 will be made available through open and controlled-access tiers, based on the type of data, its sensitivity and the specification of the contributing collection.

Data access and release under the controlled access tiers will be overseen by the CGEn Databank independent Data Access Compliance Office.

The DACO will adopt its own terms of reference (including, for instance, the frequency of its meetings, the process for the review of access applications, etc.). A DACO is typically composed of a minimum of five (5) voting members, independent from CGEn, including:

- At least one (1) scientific expert [voting member];
- At least one (1) bioinformatics researcher [voting member];
- At least one (1) expert in the technical infrastructure of the designated CGEn Databank [voting member];
- At least one (1) expert on the legal/ethical aspect of genomic research, data sharing, privacy and data protection [voting member];
- At least one (1) patient/participant representative or member of the public [voting member];
- One (1) investigator representing the CGEn Databank [non-voting member].

For certain access applications, the DACO may call upon outside experts. Such experts will not be voting members, but are instead invited to provide background expertise required to review the application.

Data management

Data access

All aspects of the Databank are overseen by a dedicated Project Manager. Databank data is made available through open and controlled access tiers (Table 1).

Table 1. Overview of CGEn databank data access tiers

TIER 1: OPEN ACCESS	
Data available (Open Access)	Data with very low risk of re-identification or sensitivity ("open data"), e.g. aggregated genomic and phenotypic data.
Who can access?	Openly accessible, through the use of APIs/CGEn website.
Mechanism	Open access
Access Process	CGEn Databank website
TIER 2: CONTROLLED-ACCESS	
Data available (Controlled Access)	Individual-level molecular and phenotypic data, including: • Phenotype data • Standard output file formats including gvcf/vcf files for SNV and indel variants, vcf files for SV and CNV calls and BAM and index files for visualization of the data
Who can access?	Canadian researchers approved by the DACO (as a clarification, researchers who have not submitted datasets can still apply to access the collection)
Mechanism	Controlled access
Access Process	Access to data by researchers requires submission of an application to the DACO. Requests must include a summary of the research project, justification for data types requested, ethics approval from the researcher's institutional REB.

Data sharing with other databases

In order to enable international data sharing and create shared resources, efforts are made to share CGEn Databank datasets with other international databases and/or biobanks. In such cases, access to CGEn Databank resources is contingent on signing of a data access agreement or collaboration agreement. Sharing of datasets or samples from other international collections with the CGEn Databank is permitted if they meet consent and regulatory requirements, and is also contingent on signing of a data access agreement or collaboration agreement.

Database closure

If any CGEn Databank were to cease activities, efforts will be made to transfer the data to a third party that agrees to comply with the CGEn databank policies and the terms of participants' consent (as a note: transfer of server storage entities or location does not constitute a transfer of data custodianship). Prior to any transfer of data custodianship and responsibilities with respect to the maintenance of the CGEn Databank, each contributing collection's Principal Investigators and local REB will be notified of such decision to transfer data and provided with the opportunity to accept the terms of transfer, or to withdraw data from collections under its purview. A decision made with regard to the transfer or closure of the database will involve the CGEn Board of Directors or other relevant bodies.

Protection of participant privacy and data security

Each participating institution has a duty of maintaining the privacy and confidentiality of their participants' nominal information, i.e., information that can be directly or indirectly traced back to an individual participant. This information is stored on secured servers in each participating collection's internal database.

To protect participants' information, upon inclusion of collection data in the CGEn Databank, data is coded according to the local collection's procedure. If the local collection's REB or institution requires data to be double-coded prior to depositing data in the CGEn Databank, double coding should is done by the collection, which is ultimately responsible for providing this new, second code to the CGEn Databank.

The only link between the participant's nominal information and the coded data stored in the CGEn Databank is the participant ID. The coded data is thereafter imported and stored in the network part of the CGEn Databank. The key to link the participant's CGEn Databank number and collection participant ID (e.g. linking log) is held locally by the collection principal investigator. The CGEn Databank does not collect or store information that directly identifies participants (e.g. names, contact information, healthcare numbers, etc.). Therefore, the possibility of obtaining nominal information on participants by consulting the shared part of the CGEn Databank is very low.

Only authorized members of the CGEn databank team, CGEn affiliated researchers, and third-party researchers approved by the DACO will have access to authorized sets of coded phenotypic and genomic data. In the case of CGEn affiliated researchers, access to datasets other than their own collection always requires DACO approval.

Management of participant withdrawal

Contributing data is voluntary and participants will continue to receive the best available care, whether or not they decide to share data with the CGEn Databank.

Participants have the right to withdraw data from a CGEn Databank at any time and without providing any reason. Withdrawal can be implemented by informing the CGEn affiliated researcher who is contributing to the CGEn Databank. Upon withdrawal from the CGEn Databank and/or the participating collection, the CGEn affiliated researcher shall contact the CGEn Databank to request withdrawal and that data be destroyed and removed from the CGEn Databank. However, data that has already been distributed for research analysis, cannot be removed or destroyed to preserve the scientific integrity of the analysis.

Intellectual property

CGEn does not claim Intellectual Property rights on the data stored in a CGEn Databank. Intellectual Property rights on derived data should not impede data usage by the researchers accessing the shared resource.