Genome BritishColumbia

Genomics positively impacts life, every day.

Genome British Columbia ("Genome BC")

Data Management and Sharing Policy

Introduction

Genome BC strategically promotes data collection, management, storage, analysis and integration as they are essential for genomic research and innovation. Our vision for data sharing is consistent with the Organisation for Economic Cooperation and Development's (OECD)1 notion that publicly funded data are public goods and as such should be openly shared to deliver economic, social and environmental benefits[2].

A vast majority of data resulting from Genome BC funded projects has value and utility beyond the purpose for which it was originally generated. Unlocking the potential of project generated data is an interdisciplinary and translational challenge, which requires the engagement of multiple participants including funders, research institutions and universities (collectively, "**Institutions**"), data owners, holders, stewards and providers, innovators, researchers, and end-users.

Genome BC's general objective is that data produced with Genome BC funding and support will be freely shared and made widely available on a timely basis, thereby enabling the release of data across the scientific community to foster innovation and growth. However, this objective needs to be balanced against any obligations owed by Genome BC to project participants, the terms of any intellectual property policies of such project participants and compliance with applicable laws and regulatory requirements, including privacy laws. Data, including the protocols and methods used to generate the data, will be shared by Genome BC and the Institutions and their researchers that are funded by Genome BC, in a manner adhering to such obligations, laws and regulations.

Applicability and Scope of Policy

This Genome BC Data Management and Sharing Policy (the "Policy") applies to:

- all new research data and the basic research, laboratory studies and protocols, field surveys, and other types of research supported by Genome BC;
- the sharing and use of research data for research, scholarly publication, educational and all other non-commercial purposes; and
- project data that transforms or links pre-existing datasets.

If Genome BC support is sought to use, develop or modify existing data, the parties seeking such funding are required to provide representations and warranties to Genome BC confirming that such parties have rights to use and/or share the existing data with Genome BC in

[1] https://www.oecd-ilibrary.org/sites/276aaca8-en/index.html?itemId=/content/publication/276aaca8-en

[2] https://www.oecd-ilibrary.org/sites/3631ee20-en/index.html?itemId=/content/component/3631ee20-en

connection with the proposed research project. Suppose there are any restrictions imposed on the use of the existing (or modified form of) data by Genome BC. In that case, the Institution(s) needs to include any such restrictions within the project proposal and research agreement.

"Intellectual Property" means any and all knowledge, know-how and/or technique or techniques, technology, trade secrets or other intellectual property, including without limitation patents and/or patent applications, invented, developed and/or acquired, or being invented, developed and/or acquired by a party as a result of the research and development activities funded by Genome BC.

The terms of this Policy are subject to any contractual obligations that may be owed, from time to time, by Genome BC to third parties under any agreements with such third parties governing GBC funded research projects, including without limitation, Institutions that may have intellectual property policies **("IP Policies")** that govern the ownership of Intellectual Property arising from research projects, and future commercialization of such Intellectual Property. In the event of any inconsistency or conflict between the terms of this Policy and the terms of such contractual obligations and/or IP Policies, as the case may be, such contractual obligations and/or IP Policies shall govern and prevail.

Data Sharing

Genome BC promotes, mandates and actively supports plans, policies and best practices for data governance including management, sharing and the provision for data access in accordance with applicable laws, regulations, standards and guidelines of the jurisdiction where the research is performed, and in a manner that removes barriers limiting access to data and other resources. To accomplish this, Genome BC promotes scientific data management and stewardship using global, academic and industry standards, including the **FAIR Guiding Principles**[3], and for data involving Indigenous people, the **First Nations Principles of OCAP®**[4], the **CARE Principles for Indigenous Data Governance**[5], and, where applicable, **Principles of Ethical Métis Research**[6], and the **Guidelines for Research Involving Inuit**[7]. **FAIR** refers to making data:

[3] https://www.go-fair.org/go-fair-initiative

[4] https://fnigc.ca/ocap-training/

[5] https://www.gida-global.org/care

[6] https://achh.ca/wp-content/uploads/2018/07/Guide_Ethics_NAHOMetisCentre.pdf

[7] https://achh.ca/wp-content/uploads/2018/07/Guide_Ethics_Inuit-Fact-Sheet.pdf

- **Findable** - Metadata and data should be easy to find for both humans and computers. Machine-readable metadata are essential for automatic discovery of datasets and services.
- **Accessible** - Once the required data has been found, the user needs to know how it can be accessed, possibly including authentication and authorization.
- **Interoperable** - The data usually needs to be integrated with other data. In addition, the data needs to interoperate with applications or workflows for analysis, storage, and processing.
- **Reusable** - The ultimate goal of FAIR is to optimize the reuse of data. To achieve this, metadata and data should be well-described so that they can be replicated and/or combined in different settings.

The core requirements of the FAIR principles which are to be followed include:

1. All relevant dataset files are described by adequate metadata.
2. All datasets are assigned a globally unique and persistent identifier that is included in the core metadata, such as a Digital Object Identifier (DOI) or Accession number.
3. The persistent identifier and adequate metadata are registered or indexed in a reliable, searchable resource (such as Genome BC's Metadata Commons).

The **CARE** principles describe how data should be treated to ensure that Indigenous governance over Indigenous peoples' data and its use are respected. CARE is the acronym for:

1. **Collective benefits** - Data ecosystems shall be designed and function in ways that enable Indigenous peoples to derive benefit from the data.
2. **Authority to control** - Indigenous peoples' rights and interests in Indigenous data must be recognized, and their authority to control such data be empowered.
3. **Responsibility** - Those working with Indigenous data have a responsibility to share how those data are used to support Indigenous peoples' self-determination and collective benefit.
4. **Ethics** - Indigenous peoples' rights and wellbeing should be the primary concern at all stages of the data life cycle and across the data ecosystem.

To effectively share data, Genome BC funded projects shall generally follow the data release and resource sharing principles of a "community resource project"[8], defined as "a research project specifically devised and implemented to create a set of data, protocols, reagents or other material whose primary utility will be as a resource for the broad scientific community." Genome BC also recognizes that in some instances the data will include proprietary information

[8] https://genomecanada.ca/wp-content/uploads/2022/05/gcdatasharingpolicies16-09-23-1.pdf

and/or know-how which may be desirable to protect, as described further below under the heading entitled "**Publication of Data Subject to IP Policies**". In addition, researchers are expected to respect Indigenous, First Nations OCAP®, Métis and/or Inuit principles regarding data and research. This may include sharing and/or returning such data and results with/to the communities involved in the collaboration.

Goals of Data Sharing

Science is based on collaboration. In order for science to function effectively so that society can benefit, science must be made open to promote meaningful collaboration and generate research-enabling resources. Data and protocol sharing, and data documentation allows scientists to facilitate the translation of research results into knowledge, products, procedures and measurable outcomes related to Genome BC's contracts and initiatives.

There are many reasons to share data. Sharing data facilitates open scientific inquiry, encourages diversity of analysis and opinion, promotes the emergence of new research, allows the testing of new or alternative hypotheses and methods of analysis, supports studies on data collection methods and metrics, and facilitates the training and development of new researchers. Additionally, it enables the exploration of topics beyond the scope of the initial studies, facilitating the creation of new datasets for new applications from the synthesis of diverse datasets, and enabling open inquiry about the impact of project activities.

The ultimate value of data is unpredictable. With advancing technologies and analytics, existing datasets may be revisited, yielding novel interpretations, and fostering innovation. The objective of data sharing is to maximize the utility and reach of research results and outputs, towards the goal of extending the contributions of each research project beyond their initial scope and providing long term value to the scientific community.

Human Subject Data

Responsible scientific data sharing practices promote both effective data stewardship and protection of privacy rights of any human research participants. Privacy rights of human subjects ("**Subjects**") who participate in any Genome BC sponsored research must be protected at all times in accordance with applicable laws. **"Personal Information"** means recorded information about an identifiable individual other than contact information, including without limitation genomic and health-related data of an identifiable individual. It is the responsibility of all Institutions and researchers involved in any Genome BC sponsored research projects to:

- protect the privacy and confidentiality rights of Subjects[9];

[9] Several resources are available for guidance on principles and best practices for protecting Subject

- only collect, use and disclose Personal Information of Subjects in accordance with the terms of, and subject to obtaining, appropriate informed consent from Subjects;
- comply with all applicable laws and regulations governing clinical trials, including without limitation maintaining the confidentiality and privacy of any Personal Information collected in accordance with applicable privacy laws; and
- meet all requirements of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans,* as well as any obligations prescribed by the relevant research ethics board.

Genome BC requires any new Subject enrollment to include informed consent from Subjects for any secondary research. If relevant, informed consent from Subjects should also be sought for data linkage with administrative health data. The responsibility for obtaining such informed consent, as well as the appropriate data access safeguards, rests with the Institutions.

All Institutions and researchers involved in Genome BC funded projects shall obtain appropriate certification and/or approval regarding use of humans, animals, radioactive materials, and/or biohazards and/or research involving possible environmental effects in the conduct of the project in accordance with applicable laws, regulations, standards and guidelines, and institutional policies including, without limitation, any approvals and/or certifications required from the appropriate research ethics board, animal care committee, or certification/safety committees, as applicable. Research involving human subjects shall meet the requirements of the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*.

Data Release and Resource Sharing Plan

Given the variety of projects that Genome BC supports, Genome BC does not prescribe the precise content for the data documentation, nor the formatting, presentation, or transport mode for data. However, Institutions and researchers must prepare and plan for data sharing, and be aware of the current state of data sharing activities and data management best practices. Applications for funding shall include a detailed data release and resource sharing plan (a "**Data Plan**") which will be reviewed by Genome BC as part of the evaluation criteria, and Genome BC reserves all rights not to fund until Genome BC approves the Data Plan. Genome BC will also monitor compliance with the approved Data Plan throughout the course of the research project. Each Data Plan shall include (without limitation) the following:

- the type(s) of data to be collected (including metadata);
- related tools, software and/or code;

---

privacy including policies for data management and sharing, and genomic data sharing (e.g., https://sharing.nih.gov/data-management-and-sharing-policy/protecting-participant-privacy-when-sharing-scientific-data)

- common structured data standards which will be applied to the data;
- use of bioinformatics and statistical analysis of the data;
- data preservation, publication, release, access and associated timelines (i.e., how soon the data will be made available and for how long);
- consideration of potential sharing of resources (by way of deposit of genomics data and corresponding metadata into open access repositories) generated by projects;
- how the management of the data will adhere to the FAIR principles;
- for data involving Indigenous people, how the management of the data will adhere to the CARE principles, First Nations OCAP®, Métis and Inuit research principles;
- factors affecting data access, distribution and reuse;
- control of the data and data governance;
- for studies involving human Subjects:
    - any obligations with regard to data handling, privacy and protection that are required by the research ethics board and/or the *Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans*;
    - collection, tracking and maintaining records of informed consent for any personal information of human Subjects being collected and used;
- deployment of reasonable physical, administrative, and technical safeguards to meet privacy and confidentiality obligations, and implementation of industry standard information security controls and security measures to protect and manage the security and integrity of data, including against potential data breaches (in compliance with applicable privacy laws, ethical guidelines and industry standards);
- specific jurisdictional, legal and/or regulatory restrictions and limitations on data sharing and/or data privacy that are relevant to the data;
- reasonable compliance with the Nagoya Protocol on Access and Benefit-Sharing (see below); and
- oversight of the data management and sharing strategy.

Examples of sample plans may be obtained online (e.g., SSHRC or NIH[10] websites). Activities in the Data Plan should also address areas such as:

- data and metadata deposition to relevant online data repositories[11] and data federations, including listing what metadata is shared on Genome BC's Metadata Commons,

---

[10] https://sharing.nih.gov/data-management-and-sharing-policy/planning-and-budgeting-for-data-management-and-sharing/writing-a-data-management-and-sharing-plan or https://sshrc-crsh.canada.ca/en/funding/opportunities/resources/guide-preparing-data-management-plan.aspx

[11] For example, GenBank (https://www.ncbi.nlm.nih.gov/genbank/) or Gene Expression Omnibus (GEO, https://www.ncbi.nlm.nih.gov/geo/)

- procedures for data documentation,
- data formatting and data exchange standards,
- software (or online data services) that conform to data format and exchange standards,
- procedures for quantifying the demand (use) for the data (i.e., the number and rate of users and records per data repository and/or data federation per year), and
- procedures for the data owner (and/or provider of the datasets) to receive feedback about the quality of the data and its access, including privacy considerations of data and metadata, especially for indigenous sovereignty, if applicable.

Familiarity with the above principles will allow Institutions and their researchers to better estimate the costs and benefits associated with their data release and resource sharing strategy, as well as their implementation plan.

When developing their data sharing strategy and subject to any obligations of Institutions under IP Policies and/or as described under "**Publication of Data Subject to IP Policies**" below, Institutions and their researchers should also consider adopting an Open Access ("**OA**") policy, the key components of which include:

1. Immediate, free online pre-publication access[12], or access upon publication with researchers retaining copyright via the **Creative Commons Attribution License**[13] (**CC BY 4.0** or equivalent).
2. Manuscripts that are posted in an OA preprint repository[14] upon submission to a journal for review (or sooner).
3. Publishing in a gold or hybrid journal[15] that offers immediate OA of the final edited version of the journal article rather than in a subscription-only journal.
4. All research outputs[16] (subject to any applicable confidentiality and/or privacy obligations) being deposited in publicly accessible, community recognized repositories, cited in the publication, and made available at the time of manuscript submission, along with any resource required to view datasets or to replicate study findings.

---

[12] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3073843/

[13] https://creativecommons.org/licenses/by/4.0/

[14] For example, bioRxiv (https://www.biorxiv.org/) or medRxiv (https://www.medrxiv.org/)

[15] Gold journals provide immediate OA for all articles, Hybrid journals provide OA for an additional fee on some articles but do not allow all of their articles to be freely accessible

[16] Such as primary data and associated metadata (with the exception of Proprietary Data and/or any Personal Information)

5. Manuscripts should be posted in an OA preprint repository upon submission to a journal for review (or sooner).

Genome BC's Metadata Commons

Researchers and Institutions are expected to deposit their project and omics related metadata into Genome BC's Metadata Commons[17] to allow other researchers to see what datasets are available and to promote further collaborations and secondary use of the generated omics data. The Metadata Commons serves as a centralized hub for storing and disseminating metadata pertaining to biological samples and omics research data. As of July 2024, usage of the Metadata Commons will be incorporated into research reports and will factor into future follow-on funding decisions.

Repositories for Data, Publications, Protocols, Lab Resources, and Software

Omics and Other Data (Including Metadata): Institutions and researchers are generally allowed to select their preferred data repository (provided that it is reasonably acceptable to Genome BC) to store project-related results. For the deposition and retrieval of genomic data, Genome BC encourages the use of internationally recognized databases (i.e., INSCD) such as the National Center for Biotechnology Information ("**NCBI**"), the European Bioinformatics Institute ("**EBI**"), and other equivalent repositories that adhere to global standards of data quality, accessibility, and ethical considerations. The data repository of the researchers' choosing must adhere to the OA criteria described above and provide long-term storage and preservation, such as those that meet the ISO's Trustworthy Digital Repository Standards[18]. In addition, they should comply with the FAIR requirements described above. For data involving Indigenous peoples, CARE, OCAP®, Métis and Inuit principles must also be followed. For data produced with Genome BC funding, Genome BC recommends that researchers describe their omics data in Data in Briefs[19], Scientific Data[20] or a similar journal.

Publications: Subject to any obligations of Institutions under IP Policies and/or as described under "**Publication of Data Subject to IP Policies**" below, Genome BC expects the following with regards to publications:

1. Articles should be published in a gold or hybrid journal that offers immediate OA of the final edited version of the journal article rather than in a subscription-only journal.

---

[17] https://metadatacommons.ca

[18] https://www.iso.org/standard/66760.html

[19] https://www.sciencedirect.com/journal/data-in-brief

[20] https://www.nature.com/sdata

2. Manuscripts should be posted in an OA preprint repository upon submission to a journal for review (or sooner).

Protocols: Experimental protocols that are employed in Genome BC funded research must be made publicly available (and updated as needed) through a protocol sharing service such as protocols.io[21]. Genome BC recommends its use for sharing standard operating procedures.

Lab Resources: All tools or reagents that are funded by Genome BC and/or result from Genome BC awarded projects must be made readily available to the community and citable as a means of supporting reproducibility and to enable further research. For assistance with citing lab resource tools, please see the Research Resource Identification Portal (RRID)[22] that aggregates information from multiple repositories and provides identifiers for citation in publications. This requirement applies to cell lines, transgenic models, plasmids/clones, antibodies, and other reagents.

Software: Software, including algorithms, scripts, and other code-based research outputs, must be made available through code repositories like GitHub[23] and assigned a DOI. If it is difficult or impossible to assign a DOI through the code repository, Zenodo[24] can also assign a DOI to code hosted elsewhere.

Genome BC believes that sharing and discussing scientific methods is crucial to achieving credible, reproducible experimental research, and so will factor use by researchers and Institutions of OA repositories, databases and sharing platforms, such as those described above, in ongoing and future follow-on funding decisions.

Compliance with the Nagoya Protocol on Access and Benefit-Sharing

Genome BC acknowledges the importance of the Nagoya Protocol on Access and Benefit-Sharing[25] as a cornerstone for the responsible utilization of genetic resources. As part of our commitment to ethical research and international biodiversity conservation goals, Genome BC

[21] https://www.protocols.io/

[22] https://scicrunch.org/resources

[23] https://github.com/

[24] https://zenodo.org/

[25] https://www.cbd.int/abs

requires all applicable research activities involving genetic resources to reasonably follow the principles and obligations outlined in the Nagoya Protocol. This may include obtaining prior informed consent from a provider country, establishing mutually agreed terms ("**MAT**") that include clear benefit-sharing arrangements, and ensuring the traceability of genetic resources throughout the research process. Researchers and Institutions must strive to ensure that their Data Plans align with the Nagoya Protocol, particularly in the following aspects:

- Access to Genetic Resources: Detailing the procedure for accessing genetic resources, including compliance with the legal and administrative measures of the provider country.
- Benefit-Sharing Agreements: Describing the measures in place for fair and equitable sharing of benefits arising from the use of genetic resources, which could encompass monetary and non-monetary benefits such as technology transfer, capacity building, and contribution to the local economies where the genetic resources were sourced.
- Monitoring and Compliance: Outlining the mechanisms for monitoring the utilization of genetic resources and compliance with MAT, ensuring transparent and lawful international collaboration.

Publication of Data subject to IP Policies

Genome BC recognizes that certain Institutions may develop, own and wish to protect and commercialize Intellectual Property (which may include patents and proprietary data) in accordance with the terms of their respective IP Policies. Project data generated from such projects will be published, and protected, in accordance with the applicable IP Polic(ies) of such Institutions. Institutions will ensure that data within patent filings is normally made public through either the pending publication that precipitated the filing, or jurisdictional publication of the patent application itself. Genome BC understands that an Institution's desire to protect its Intellectual Property may justify a need to delay disclosure of research findings. Review and approval of any proposed publication(s) will be conducted in accordance with the Institution's publication policies. Institutions and researchers shall refrain from submitting any proposed presentation or proposed publication until the filing of one or more patent applications with one or more patent offices directed to such patentable subject matter, or the expiration of six (6) months, whichever is earlier, after which the applicable Institution and researchers may proceed with the submission of the proposed presentation or proposed publication.

Genome BC expects that Institutions will develop Intellectual Property, licenses and material sharing policies that will promote commercialization and product development without compromising the continuing availability of new research tools to the scientific community and industry. Researchers and institutions shall use commercially reasonable efforts to ensure that Intellectual Property created or developed by Genome BC funded projects is commercialized in a way that maximizes benefits for BC and Canada. Institutions will endeavour to publish and release data for the furtherance of common good in pursuing the furtherance of scientific

knowledge through public communication, as a requirement of public funding while protecting any Intellectual Property and respecting IP Policies.

Timeliness of Data Sharing

Recognizing that the value of data often depends on their timeliness, data sharing and publication should occur in a timely fashion. Genome BC expects the timely release and sharing of data on an OA basis to be no later than six (6) months after the original publication date of the main findings from the final dataset generated for a project. The specific time will be determined by the nature of the data collected. Data from small studies can typically be analyzed and submitted for publication relatively quickly. If data from large epidemiologic or longitudinal studies are collected over several discrete time periods or phases, it is reasonable to expect that the data will be released in phases as they become available or as the main findings from a research phase are published.

Funding Compliance with the Data Plan

Genome BC recognizes that it takes time and resources to manage data appropriately and prepare it for sharing. Thus, applicants can request funds from Genome BC to support such data management and sharing in accordance with the terms of a Data Plan, including publication as described above, in their project application and proposal. Institutions who prepare a detailed Data Plan and incorporate data sharing within their initial design of the research project may more readily establish adequate procedures for data management and protection, and share a useful dataset with appropriate documentation.

Funding Recognition

The financial contribution and support of Genome BC should be recognized and attributed for any research publications, activities and projects that Genome BC funds, in the following:

- the repository records where data and protocols are deposited
- in the acknowledgement sections of journal articles
- in any related public communications (including but not limited to web sites, publications, news releases, presentations, annual reports, and on-site signage), publications, press releases and promotional material

GBC's funding support shall be recognized using the following written acknowledgement: *"This research was supported by Genome British Columbia [project code]."* In addition, visual presentations such as seminars, poster presentations and websites must include the Genome BC logo.