GenomeCanada

Data Release and Resource Sharing

POLICY PRINCIPLE

Genome Canada is committed to the principle of rapid data release and sharing of unique resources to the scientific community; Genome Canada-funded projects must therefore share data and resources in a timely fashion with minimal or no restrictions. By providing the scientific community with timely access to the outputs of Genome Canada-funded projects, this data and resource sharing policy is intended to accelerate the translation of research for the benefit of humankind.

This policy was originally established by Genome Canada's Science and Industry Advisory Committee and approved by Genome Canada's Board of Directors on June 17, 2005. The revised policy was approved by Genome Canada's Board of Directors on September 18, 2008 for application to all projects funded by Genome Canada after September 1, 2008.

MECHANISMS FOR DATA AND RESOURCE SHARING

Genome Canada expects researchers to share data and resources as rapidly as possible. Where the goal of the project is to produce data or resources for the wider scientific community the project must follow the data release and resource sharing principles of a "Community Resource Project", defined as "a research project specifically devised and implemented to create a set of data, reagents or other material whose primary utility will be as a resource for the broad scientific community." This definition and the associated data release and resource sharing principles were developed at a meeting held in January 2003 in Fort Lauderdale[1]. Genome Canada encourages the application of the principles of rapid, pre-publication data release to other types of projects and is working with other research funders to promote this practice.

Genome Canada recognizes publication as a vehicle for data release, and, at a minimum, expects data to be released and shared no later than the original publication date of the main findings from any datasets generated by that project. For large datasets that are collected over several discrete time periods or phases, it is reasonable to expect that the data be released in phases as they become available or as main findings from a research phase are published. However, at the conclusion of a project, all data must be released without restriction.

Genome Canada recognizes the need to protect patentable and other proprietary data; however data must be released no later than the date that patents (including provisional patents) have been filed or at the time of publication, whichever comes first. A maximum extension delay of 90 days is acceptable if there are extenuating circumstances. Permission to delay release of data beyond this timeframe must be obtained in writing by Genome Canada, which will use independent advisors to evaluate the request when appropriate. In addition, non-disclosure agreements between sponsors and researchers should not be unduly restrictive and must allow publications within a defined and reasonable period of time.

Genome Canada approved projects must also address the sharing of resources generated by the projects such as unique biological specimens and computer programs designed to analyze datasets. Biological reagents such as unique strains should be deposited into repositories such as ATCC and computer programs designed to analyze large datasets should be made available to others through the use of license agreements that adhere to "open source" principles (see for example, http://www.opensource.org/).

[1] The definition of "community resource project" was developed at a meeting held on January 14-15, 2003 Fort Lauderdale, Florida. The report on the conclusions of the meeting, "Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility", can be found at http://www.wellcome.ac.uk/stellent/groups/corporatesite/@policy_communications/documents/web_document/wtd003207.pdf

DATA AND RESOURCE SHARING PLANS

Genome Canada applicants must provide a Data and Resource Sharing Plan as part of their application. This plan must address the following issues:

- **Data and resource types** – what data and resources will be generated;
- **Timing and mechanism for sharing** – for each data and resource type, when, how and where will these be made available. Where there are recognised public databases and repositories these must be used, and if none are currently available, what are the plans for making the resource in question available to the community at large;
- **Quality** – what quality control/assurance mechanisms will be in place;
- **Standards** – are there community standards for the data and/or resources being generated, how will the project conform to these. Genome Canada expects that data and resources generated will conform with internationally accepted standards, and reference to these standards should be made when these are available;
- **Ethical, privacy and confidentiality issues** – if the data could be of a potentially sensitive nature, how will this be handled? Where the research involves human subjects, how will the interests of the research participants be protected? How does the Data and Resource Sharing Plan comply with the terms of the consent;
- **Intellectual Property (IP)** - will there be any restrictions or delays on data and/or resource sharing to ensure protection of any IP or proprietary data and/or resources[2];
- **Terms and Conditions -** what terms and conditions, if any, of access and use of the data and/or resources will be implemented. Please note, when making data and resources available, researchers cannot place limits on questions posed or methods used, nor require co-authorship as a condition for receiving data or resources.

Genome Canada and its review committees will review the applicant's proposed data and resource sharing plan to verify that it conforms to the Genome Canada policy and funds will not flow until an acceptable plan has been approved and incorporated into the terms of award. Genome Canada and Genome Centre staff will monitor adherence to this policy by funded researchers through a variety of mechanisms including interim review.

USE OF DATA AND RESOURCES

Genome Canada believes that the scientific community benefits from the practice of rapid sharing of data and resources. In order to ensure that this practice continues and to respect the interests of the data and/or resource generators, Genome Canada expects that users of data and resources will acknowledge the source and abide by any terms and conditions of use.

Genome Canada will keep this policy under review and continue to work with other research funders to promote best practice in this area.

[2] Please see Genome Canada's Intellectual Property Policy: http://www.genomecanada.ca/medias/PDF/EN/IntellectualProperty.pdf

EXAMPLES OF INTERNATIONAL DATA AND RESOURCE REPOSITORIES

The following table provides examples of where various data types or unique resources produced by Genome Canada-funded projects may be deposited. The examples given are some of the recognized databases and repositories currently available for various data and resource types; this is not intended to be an exhaustive list.

Data/Resource type	Database	URL for submission
DNA, RNA and Protein Sequences	DDBJ/EMBL/GenBank[1]	http://www.ncbi.nlm.nih.gov/Genbank/submit.html
EST sequences	DDBJ/EMBL/GenBank	http://www.ncbi.nlm.nih.gov/dbEST/how_to_submit.html
STS sequences	DDBJ/EMBL/GenBank	http://www.ncbi.nlm.nih.gov/dbSTS/how_to_submit.html
Microarray data (eg. gene expression and CNV)	GEO	http://www.ncbi.nlm.nih.gov/geo/
SNPs	dbSNP	http://www.ncbi.nlm.nih.gov/SNP/
Human genotype and phenotype (GWAS)	dbGaP	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gap
Trace and Short Reads from Next Generation Sequencers	Trace Archive and Short Read Archive	http://www.ncbi.nlm.nih.gov/Traces/trace.cgi?cmd=show&f=rfc&m=main&s=rfc And http://www.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?cmd=show&f=main&m=main&s=main
Protein Structures	PDB	http://www.wwpdb.org/
Interactions	IntAct	http://www.ebi.ac.uk/intact
Human gene nomenclature	HGNC	http://www.gene.ucl.ac.uk/nomenclature
Mouse gene nomenclature and mouse phenotype and expression data	MGI	http://www.informatics.jax.org/mgihome/nomen/index.shtml http://www.informatics.jax.org/submit.shtml
Biological strains	ATCC	http://www.atcc.org/
Software Code	SourceForge.net/ Canadian Bioinformatics Help Desk Software Repository	http://sourceforge.net/ http://www.gchelpdesk.ualberta.ca/repository/contents.php

[1] DDBJ, EMBL, and GenBank are equivalent databases, all part of the International Nucleotide Sequence Databases Collaboration, and submitting to any of these is the same as submitting to GenBank. It is most convenient for Canadian scientists to submit to GenBank, but scientists are free to submit to any of the three databases