



Cristiana Paşca Palmer, PhD  
Executive Secretary  
Secretariat of the Convention on Biological Diversity  
United Nations Environment Programme  
E-mail: [secretariat@cbd.int](mailto:secretariat@cbd.int)

May 29, 2019

Dear Dr. Paşca Palmer,

In response to notification 2019-012 “Digital Sequence Information on Genetic Resources: Submission of views and information and call for expression of interest to undertake studies”, Canada is pleased to provide its views below regarding concepts, and examples of benefit-sharing arrangements relating to such sequence information.

Best regards,  
Basile

CBD National Focal Point

**Basile van Havre**

Director General for the Biodiversity Policy and  
Partnerships Directorate  
Canadian Wildlife Service  
Environment and Climate Change Canada  
351 St-Joseph  
Gatineau, QC  
K1A 0H3  
Canada  
Telephone +1 819 938 3935

Directeur Général des politiques et des partenariats  
sur la biodiversité  
Service Canadien de la Faune  
Environnement et Changement Climatique Canada  
351 St-Joseph  
Gatineau, QC  
K1A 0H3  
Canada  
Téléphone +1 819 938 39



## CANADIAN INPUT TO NOTIFICATION 2019-012

### (a) Views and information to clarify the concept of digital sequence information on genetic resources

Canada supports discussion on the relationship between “digital sequence information on genetic resources” (DSI) and the objectives of the Convention, within the CBD, and recognises that “DSI” is an increasingly vital tool for the implementation of the Convention, contributing to Articles 7, 8, 9, 10, 12 and 17. While DSI is uncovered/generated by research conducted on genetic resources, Canada takes a firm position in many fora (CBD, ITPGRFA, CGRFA, WHO) that DSI is not equivalent to tangible genetic resources or genetic material, and continues to resist language that implies any such equivalency.

Regarding terminology, Canada is uncomfortable with the use of the term “digital sequence information on genetic resources”, which was introduced by CBD decisions and not, until these discussions, used by the scientific community. In summary, we consider that it is ill-defined; too general/broad to be operationally meaningful; and it can have a variety of interpretations under different contexts.

It is unclear as to what “DSI” is intended to refer. For example, does it solely refer to complete biomolecule strings, or also to partial strings? Does it solely refer to hereditary biomolecule strings, or also non-hereditary strings (such as amino acid strings)? To naturally-occurring biomolecule strings, or also non-natural biomolecules (e.g. lab-modified biomolecules; hypothetical biomolecules (lab artifacts, instrument errors); recent non-biological computer data sets (binary codes) being stored as DNA for efficient digital information storage, aka “DNA digital data storage”)?

Unpacking the term “digital sequence information on genetic resources”, we note:

- i) Regarding “digital”, the use of “digital” implies that these resources are no longer in their initial biomolecular format; further, the fact that genetic sequences can be conveyed digitally is of secondary importance. The “digital” aspect only refers to a particular mode of storage or mode of transmission and such sequences listed on a sheet of paper would still be genetic sequences.
- ii) Regarding “sequence”, the sequences in question are biomolecular sequences, not (for example) mathematical sequences or sequences of events, so there needs to be a descriptor for the type of sequence.
- iii) Regarding “information”, the term is problematic as a term in this context:
  - “information” is generally achieved after data collection and analysis;
  - various degrees of “information” can be achieved depending on when/how much/the way data is processed,
  - “information” outputs can change as new analyses are performed;
  - while original data stay the same, “information” can be outputs on a continuum such as discrete packets (genes) or broader aggregates (entire metabolic pathways);
  - how (i.e. at what state of data processing) “information” is generated usually needs to be specified to be meaningful.



- iv) Regarding “genetic resources”, recognizing that this term continues to provoke debate, Canada firmly holds nonetheless that the negotiated and agreed CBD definition applies to tangible matter, and the definition’s wording “containing functional units of heredity” clearly points to the involvement of DNA and RNA rather than other types of biomolecules (such as amino acids).

We note also that although the terms “genetics” and “genomics” are sometimes used interchangeably, they have different meanings: genetics is the science of inheritance and genomics is generally considered to be study of a complex set of genes, their expression, and how they interact with other genes and their environment.

The consideration of public health approaches to this issue has potential to contribute clarity to the CBD discussions, because word meanings and terminology must be explicit for operational public health and biomedical science exchange. “Genetic sequence data” is the term employed by the WHO PIP Framework and defined in a relatively clear and practical manner as “the order of nucleotides found in a molecule of DNA or RNA... contain[ing] the genetic information that determines the biological characteristics of an organism or a virus”.

Operationally, public health practitioners and biomedical science tend to use more granular / operational language to refer specifically to the type of “digital information”. For example:

- *Hereditary* genetic sequences can be acquired in part (‘partial’ sequences) or to completion (‘complete’ sequences). Is the sequence a subcomponent sequence (e.g. sequence for a gene? A metabolic pathway? Or a chromosome?) vs. Is the hereditary sequence for an entire organism, namely its intact genome?
- *Controlled Vocabulary* terms should also be used to fully describe the level of *quality* for each genetic sequence data resource. Operationally, this is helpful for conveying the degree of data accuracy, level of curation and the type of curation. For example, when describing a *genome* resource, there are important QUALIFYING terms applied, such as:
  - “standard draft” (describes the minimum information needed for submission to a public database); “high quality draft” (describes sequences with little to no manual review); “improved high quality draft” (describes data, which has undergone review by people or by machines to some extent to indicate that most of the genetic data is assembled correctly, but some errors may still be present); “annotation-directed improvement” (in which genetic information in various gene regions is represented as accurately as possible); “non-contiguous finished” (includes sequences that have been reviewed by both people and machines and is considered complete except for “recalcitrant regions” that are proving problematic for genome closure); “finished” (which describes un-gapped complete genome sequences that have minimal errors, if any<sup>1</sup>).

Canada thus prefers the terms “genetic sequence data” or “genomic sequence data”, both of which, from a science operations viewpoint, better describe what “is” the sequence, and the codified and

---

<sup>1</sup> Source: [The Genomics Standards Consortium](#)



transmissible nature of that sequence. There should also be qualifying/subheading statements such as “partial”, “complete” sequence, etc. We shall use this term “genetic sequence data” (GSD) in the rest of the document.

Regarding domestic ABS measures with relevance to GSD, at the federal level, Canada does not have laws or regulations specifically targeting ABS, but the Government of Canada is committed to openness, transparency and information sharing through the Open Government Partnership. Since joining the Open Government Partnership, a multilateral initiative to foster greater transparency and accountability, Canada has developed three national action plans aimed at enhanced access to information and expanded open data, among other things. The Ministry of Environment and Climate Change Canada is the federal lead on the Open Science element of Canada’s 2016-2018 Action Plan on Open Government.

(b) Views and information on benefit-sharing arrangements from commercial and non-commercial use of digital sequence information on genetic resources

As GSD is effectively obtained via the utilization (the conducting of research...) on genetic resources, it should be regarded as a result of such utilization. The sharing of these results (the GSD) via access to databases and technology should thus be regarded as a valuable form of non-monetary benefit-sharing, shared openly and globally or via benefit-sharing arrangements as set out in mutually-agreed terms to cover the tangible genetic resources.

Canada is committed to open data and information sharing, but the extent to which such open sharing is, or can be, feasibly tracked – for example to identify and/or analyze which entities are accessing GSD and are thus beneficiaries of such sharing – varies according to the purpose and resources of the databases or institutions concerned. Data tracking/tracing mechanisms are more of a foundational consideration for public health laboratory management systems, for example, than natural history collection management systems.

A few examples follow of Canadian institutions and projects that generate and share GSD for global conservation, sustainable use and public health-related benefit:

The Barcode of Life Data System (BOLD; <https://boldsystems.org>) is an open-access cloud-based data storage and analysis platform designed to support the assembly and use of DNA barcode data, developed at and coordinated by the Centre for Biodiversity Genomics (CBG) at the University of Guelph. BOLD is heavily-used, evidenced by the 7M+ barcodes that it hosts, representing 600K+ species, and its 25,806 users in 144 countries. Although data sharing and collaboration are primary BOLD goals, the platform adopted stringent security policies to ensure the privacy of its users. Information contained in BOLD data records (including DNA barcode sequences) is not generally disclosed through BOLD unless the corresponding project or dataset is published by its contributor. Sequence data contained in BOLD is used by the BOLD Identification Engine to provide DNA-based taxonomic identifications to public users, and provenance data and images submitted to BOLD become publicly visible through the public BOLD Taxonomy Browser. BOLD regularly submits record updates to GenBank (which are mirrored on the other databases in the International Nucleotide Sequence Database Collaboration, INSDC); while records submitted remain private on GenBank for one year to allow time for submitters to publish their findings first, BOLD encourages its users to initiate pre-publication release of their data. As well as hosting BOLD,



CBG serves as the Secretariat for the International Barcode of Life Consortium, a research alliance of 30+ nations that aims to deliver a DNA-based identification system for all animals, plants and fungi as part of the long-term goal of establishing an Earth observation system for biodiversity (<https://ibol.org>). As a provider of sequencing services, CBG itself uses an MTA to clarify ownership, use and storage of genetic resources and GSD it handles. Through a series of collaborative initiatives with the CBD Secretariat and Canada-Americas Trade-Related Technical Assistance Program, and in partnerships with research institutions, government agencies and other organizations worldwide, CBG has facilitated the dissemination of scientific and technical advances in DNA barcoding, partly through capacity-building events in developing countries and training courses at CBG.

Canada is a voting participant of the Global Biodiversity Information Facility (GBIF). The Canadian Museum of Nature (CMN), a part of the Governing Board of GBIF since its earliest days, is a strong supporter of freely distributed biodiversity data. Over 30% of its specimen information, from botanical, zoological and palaeobiological collections from Canada (75%) and other countries, is now in digital format (>800K data points) and its goal is to completely digitize key information about its specimen holdings. It regularly updates its electronic specimen data at the GBIF portal – last year over 1B CMN data points were downloaded freely through GBIF – and intends to do so with GSD related to its specimens when this kind of data is accommodated by GBIF's distributed system in future. The CMN has a DNA sequencing facility on site, purchases sequencing services for partial or complete genome sequences that are related to its collections, and contributes GSD via BOLD. The CMN's own research is non-commercial and all loans are for non-commercial scientific research or education purposes; much of the CMN's research can be applied to conservation practices. Any process that restricts current museum practices of loaning specimens and sharing data, including information about genetic sequences, will seriously disrupt the scope of specimen-based research and international collaboration.

Canada's federal government Genomics Research and Development Initiative (GRDI) supports genomics research in eight federal science departments and agencies. As well as supporting development and maintenance of Canadian agricultural and public health genomics, GRDI also supports key environmental programs such as the EcoBiomics project, co-led by Agriculture and Agri-Food Canada (AAFC) and Environment and Climate Change Canada. This project uses metagenomics approaches to develop new knowledge to improve water quality and soil health by comprehensively characterizing aquatic microbiomes, soil microbiomes and invertebrate zoobiomes in varying habitats, and testing hypotheses to enhance environmental monitoring, assessment and remediation activities. It aims to establish comprehensive baselines for assessing future changes to water and soil biodiversity at key long-term environmental monitoring sites in Canada. It uses known GSD from sources such as GenBank and BOLD to identify the taxa in the samples and potentially their role in the ecosystem being studied, but also contributes to the identification and naming of taxa that are not known to science, and shares this knowledge openly. Such biomonitoring is essential to understand taxa and communities that may need conserving and thus potential regulatory action. There is also a strong economic tie to the production systems involved, as maintaining diversity is typically correlated to healthy systems and thus production, as well as monitoring for alien invasive species and functional changes that could damage productivity. The EcoBiomics project both benefits from and contributes to global GSD-sharing. Without GSD sharing, reference sequences will be internalized and there will be essentially no libraries available



to carry out identification and quality control. This would obstruct biodiversity research at national and international levels, especially on microbial and invertebrate life, and have a ripple effect to industry in Canada and elsewhere, and the production systems they serve. Such projects would no longer be possible if access to reference libraries of GSD and associated annotations becomes complicated and expensive.

AAFC also conducts research to document pollinator diversity in our agricultural landscapes and studying the impacts of land use, pesticides and pathogens on bee populations. This work involves sampling large numbers of bees, which must all be identified to species level. For such identification, AAFC often sequences various genomic regions of these bees and compares the sequences to those that have been made available in online databases through collaborative efforts of the scientific community. Access to GSD of bees from other countries is vital to be able to detect any newly introduced species; researchers also make use of worldwide GSD of pathogens known to negatively impact bees in other parts of the world, in order to screen bees sampled in Canada. The GSD from the thousands of bees and pathogens that we have produced and shared are invaluable tools for other researchers worldwide.

GSD allows for rapid distinguishing of species that are potentially invasive and harmful from those that are beneficial and part of natural ecosystems. Pesticides are often the primary short-term solution for crop protection against invasive insects. However, alternative pest management solutions are urgently needed as public concern over pesticide use continues to rise, and as many pesticide chemistries are phased out in Canada. Science-Based Departments and Agencies funded by GRDI, as well as partners such as CABI, conduct research to identify organisms from all life stages through the development of new genomics tools for efficient detection of quarantine and invasive species, thus substantially bolstering Canada's operational strategy for both prevention and effective eradication or mitigation of new invaders. Through the Quarantine and Invasive Species Project (2016-2011), 27 standard operating procedures were developed for DNA extraction for DNA barcoding, partial genome and whole genome sequencing of specific quarantine and invasive species and were transferred to the Canadian Food Inspection Agency for diagnostic purposes. This project is central for federal agencies working on invasive alien species, as it provides a diagnostic tool for the early detection (significantly lowering risk of establishment), surveillance and management of thousands of species. GSD and other information is shared openly and publicly via BOLD, GenBank and GBIF, and AAFC is developing a biodiversity portal to facilitate the open sharing of these reference sequences.

The identification of plant specimens (and, specifically, of wood and forest products) in international trade depends on access to GSD and the effective international sharing of such information. This includes the identification of regulated species, of regulated geographic populations of species, and of individual specimens where legal acquisition is in doubt. Such information also provides forensic evidence to support enforcement of international agreements that regulate trade in wildlife specimens, such as CITES. Currently, Natural Resources Canada – Canadian Forestry Service researchers are using genomic marker data from well-characterized forest tree species to develop timber identification methods that will contribute to international efforts, including by the Global Timber Tracking Network (GTTN; [globaltimbertrackingnetwork.org](http://globaltimbertrackingnetwork.org)) to address illegal harvesting and trade of forest products. The



voluntary exchange of GSD (and of vouchered samples required to verify taxonomic accuracy) through international cooperation between scientific institutions (herbariums, xylaria, etc.) is, therefore, a constructive tool that can be used to support and preserve biodiversity in countries of origin.

Present day biomedical and public health bodies are actively discussing how to make GSD more openly available to benefit humankind. This is important because nucleotide sequences contain crucial information about the structure and specific properties of an organism, including virulence or transmissibility characteristics. Whole genome sequencing (WGS) is modernizing foodborne disease surveillance, providing more accurate, higher resolution information than earlier methods. Since 2017, Canada has been generating and using GSD for routine surveillance, outbreak detection and outbreak response via the PulseNet Canada network, a data-sharing collaboration of Canadian public health and food safety agencies and labs. GSD is generated by the National Microbiology Laboratory (NML), the Canadian Food Inspection Agency (CFIA), and by Provincial Public Health Laboratories, and routinely shared among all Federal, Provincial, and Territorial partners involved in foodborne disease surveillance and outbreak response (laboratory and epidemiology), including Health Canada, and the CFIA. Data sharing has been supported and flowing in this way since 2004, when Pulsed-Field Gel Electrophoresis (PFGE) was used rather than GSD. As of 2019, GSD generated from foodborne disease surveillance in Canada will be made publicly available (with proper protections of patient privacy) on the INSDC. This is a result of several years of collaborative work with NML and all provincial public health laboratories to develop best practices and procedures to make GSD publically available, greatly increasing the value-added of these data for broader public health and food safety study. PulseNet Canada is part of the larger PulseNet International community, which works together to share data, standardize surveillance tools, and build capacity in 88 countries<sup>2</sup>.

Sharing GSD has provided benefits to the wider research community leading to more expertise to solve public health problems. For example, through a GRDI-funded collaboration between federal and academic researchers, the Food and Water Safety project aimed to capture the genomic diversity of pathogenic E. coli from human, food, animal and environmental sources in Canada from 1980 to 2013. The generated GSD was made publically available through the INSDC and provides a huge resource of the full One Health spectrum of disease transmission. Generating and sharing GSD was a critical component of the 2010 Haiti Cholera outbreak investigation in partnership with the US Center for Disease Control (CDC). This was a high profile international instance where the power of WGS was used to inform the source of the outbreak. WGS data from this investigation was shared with the CDC and other international collaborators and deposited into INSDC.

In conclusion, Canada views GSD-sharing as an important form of non-monetary benefit-sharing. The various applications of globally-shared GSD, such as those included in this submission, make increasingly critical contributions to the implementation of the Convention at national and international levels, as well as other international agreements, bringing benefits to national and global society. While providers

---

<sup>2</sup> [PulseNet International: Vision for the implementation of whole genome sequencing \(WGS\) for global food-borne disease surveillance.](#)



---

and users may negotiate mutually-agreed terms for the utilisation of genetic resources accessed according to domestic ABS measures, and these terms may set out how benefits, including GSD, resulting from such utilisation may be used and shared, Canada strongly supports the free exchange of GSD. It is Canada's view that rather than restricting access to GSD, emphasis and effort should be expended on bioinformatics collaboration and capacity-building to better enable all countries to use globally-shared GSD in support of their national conservation and sustainable use priorities and obligations.