



Digital sequence information on genetic resources – concept and benefit-sharing

Preamble

At the 2018 Conference of the Parties to the CBD and Meeting of the Parties to the Nagoya Protocol in Egypt decisions were adopted to gather information on Digital Sequence Information (DSI) in order to inform further discussion by the Parties. The decision requests Parties, other Governments, indigenous peoples and local communities, relevant stakeholders and organizations to submit their views and information:

- (a) To clarify the concept, including relevant terminology and scope, of digital sequence information on genetic resources and if and how domestic measures on access and benefit-sharing consider digital sequence information on genetic resources;
- (b) On benefit-sharing arrangements from commercial and non-commercial use of digital sequence information on genetic resources.

The areas of research that use “DSI” are characterised by rapid growth and innovation. The research carried out by CETAF members in the realms of taxonomy, systematics, description, environmental and biodiversity studies, evolutionary development, conservation support etc. are no different in this respect. Terminology applied to “DSI” and modalities of its use when agreed need to be unambiguous and ‘future-proofed’ to whatever extent possible. This will provide certainty and a firm base to carry on research and benefit sharing.

This consideration will affect the characterisation of “DSI”, and results in our proposal below.

Currently, the vast majority of research results from all countries are made publicly available, as are the data used in scientific studies, which includes “DSI”. The ability to replicate other researchers’ results is a cornerstone of all scientific research, and this requires the data to be available. That “DSI” is freely available has raised concerns among some developing countries that it could be used in commercial applications without triggering an obligation to share benefits with the provider countries from which the original genetic resources were acquired. CETAF members understand this concern (although this does not apply to the vast majority of research carried out by CETAF members, which is non-commercial), but also wish to highlight the large number of non-monetary benefits which are shared by the scientific community which actively contribute to and support the objectives of the CBD. These benefits were discussed in our submission to the Executive Secretary in 2017¹ and will not be repeated here.

Research data including “DSI”, when published, are maintained to the standardised quality norms of the global research community and available for use in Provider and User Countries at zero marginal cost. This important function of science is reflected in Article 15 of the CBD, which calls Parties to take *legislative, administrative or policy measures ... with the aim of sharing in a fair and equitable way the results of research and development and the benefits arising from the commercial and other utilization of genetic resources with the Contracting Party providing such resources.*

CETAF is committed to benefit-sharing and has developed best practices for its members to facilitate this. For “DSI”, the ABS Core Group believe that the most effective basis for benefit-sharing is on a

¹ <https://www.cbd.int/abs/DSI-views/CETAF-DSI.pdf>



global basis for the common good, in the manner required by Aichi Target 19. This will require a functional common set of technical standards. Promising practical examples demonstrating such standards are available, including the International Nucleotide Sequence Data Collaboration (INSDC), Global Biodiversity Information Facility (GBIF) and the Barcode of Life Database (BOLD), which address both technical and legal issues, in the case of GBIF explicitly operating within an intellectual property rights framework. We think building on these should be preferred instead of implementing new systems that aim to restrict and regulate “DSI”. Because of the exponentially growing amount of data, developing new systems with additional legal and policy-related requirements would be a difficult and expensive task with unknown results but potentially negative impact on science globally, and particularly on CBD implementation.

This submission builds on the 2017 submission from CETAF to the Executive Secretary on Digital Sequence Information on genetic resources – benefits of their use and their public availability for the three objectives of the Convention on Biological Diversity, and ramifications of restricting access to DSI².

Contents

Comments in Detail

1. The concept, including relevant terminology and scope, of “digital sequence information” on genetic resources
2. Domestic measures on access and benefit-sharing considering “digital sequence information” on genetic resources;
3. Benefit-sharing arrangements from non-commercial use of “digital sequence information” on genetic resources.

Summary

Comments in detail

1. The concept, including relevant terminology and scope, of “digital sequence information” on genetic resources

The term “Digital Sequence Information” is not used by CETAF members in their work and appears to be limited to policy discussions³. The very diverging interpretations of the term now current and the resultant ambiguity make it important to use a different term of precise meaning.

One issue lies in the term ‘information’. We understand that while ‘data’ are observations of naturally occurring states lacking extrapolated meaning, ‘information’ emerges through cognitive (or other) processing and application of data. In the “DSI” context ‘data’ are extracted from naturally-occurring genetic resources – the genetic resource itself when accessed has no intrinsic ‘information’.

² <https://www.cbd.int/abs/DSI-views/CETAF-DSI.pdf>

³ Laird & Wynberg, 2018, A Fact-Finding and Scoping Study on Digital Sequence Information on Genetic Resources in the Context of the Convention on Biological Diversity and the Nagoya Protocol. CBD/DSI/AHTEG/2018/1/3



The relevant data are the arrangement of nucleotides on strands of naturally occurring DNA or RNA⁴. This is 'Nucleotide Sequence Data' (NSD). Information about the genetic resource arises through analysis of these data.

Free sharing of both data and information, so they are available to researchers and other users in Providing and other Countries, is the usual practice in the non-commercial sector, and we view such provision as constituting a shared benefit. However, there are important differences between the two.

In practice, since 'information' is developed through analysis of 'data', it is potentially covered by Intellectual Property Rights (IPR). Sharing of IPR is identified as an example of a non-monetary benefit in the Nagoya Protocol Annex; consequently, it can be seen that IPR are a result of research and not under sovereign rights of a country as are natural resources (genetic resources).

Analysis of NSD is a key tool for taxonomy. The data used for analysis are aggregated from naturally occurring GR and downloaded from databases such as GenBank and the other INSDC⁵ members' databases, and include non-coding and coding sequences, regulatory sequences, conserved sequences, genes that encode specific traits, and 'junk' DNA⁶; sequences do not need to have a known function to be of relevance in our research. There is no maximum size for a usable sequence. Analysis might be of single genes, multiple genes, entire genomes of organisms, of a clade (pangenome) or environmental samples (metagenomes). The results of analysis are interpreted to further our understanding of biological diversity.

We propose that discussion on "DSI" distinguishes between data (NSD), and information. We also propose that the concept of "DSI" be restricted to NSD. We note that this clear concept is also in line with 'Genetic Sequence Data' (GSD) as proposed by some Parties to the CBD. We specifically exclude 'digital' from our proposed terminology to avoid inappropriate restriction to a single current means of data storage and transmission of aggregated data from GR.

For the vast majority of taxonomic and systematic research, it is not the discovery and application of functions of the genes *per se* that are important. In any case, since the function can only be discovered by experiment or through the application of existing knowledge to predict or test for function (i.e. extrapolation of data including through automatic means) this would come under the heading of information.

Sequence data may be associated with a set of other data to increase its scientific value, such as:

- i) Collection site of the organism or sample from which the NSD was obtained;
- ii) the date on which it was collected;
- iii) the name of the collector;

⁴ Nucleotides are the subunits that are connected into long chains to make nucleic acids (DNA and RNA). The four types of nucleotides in DNA are Adenine, Thymine, Guanine, and Cytosine, and in RNA Thymine is replaced by Uracil. The five nucleotides are usually abbreviated to A, T, G, C and U. The order in which these nucleotides occur in a strand of DNA or RNA is the DNA or RNA sequence or Nucleotide Sequence.

⁵ International Nucleotide Sequence Database Consortium

⁶ Sequences currently of no known function



- iv) the place where a physical voucher is stored (if it is retained) and the unique identifier of that voucher;
- v) the taxonomic name of the organism from which the DNA was sequenced.

While we do not consider this to be NSD, it is helpful contextual information and, where appropriate (and where it exists) it can be made available with NSD to which it applies. Associating these data is scientific best practice, but far from all sequences stored in public databases are associated with all of these data. Permit conditions may be stored as part of the record.

2. Domestic measures on access and benefit-sharing considering digital sequence information on genetic resources

We are aware that some countries are introducing domestic legislation including NSD. This gives us cause for concern. One is terminology; countries are using a variety of terms that are sometimes of unclear meaning, are often used inconsistently and thus increase legal uncertainty, instead of reducing it. Another is practical. CETAF members expect to seek permits or their equivalent when collecting biological specimens for research or addition to collections, and to share agreed benefits. However, these are negotiated and set out in bilateral arrangements, made in the context of significant investment of time and resources to establish scientific work in one country. Use of NSD includes work that is very different. Often the research is not targeted at a particular country but at a particular species or group of species (perhaps hundreds or even thousands of species). The uses to which the NSD are put include for example simple comparison to identify an organism or taxon from which a similar sequence has been isolated (or such a sequence within an eDNA sample, for example), developing hypotheses of evolutionary relationships and assessment of biodiversity richness. In the overwhelming majority of cases in our sector the outputs are information-based and shared globally. If countries require bilateral agreements before NSD are used this is likely to have several detrimental outcomes. Some researchers may simply ignore the requirement, exacerbating mistrust and leading to breakdown of both relationships and the growing incorporation of ABS in research workflows globally. Others will develop algorithms to avoid use of NSD from countries making such requirements, leading to fewer data and less information being generated to address biodiversity management priorities. Such a limitation would compromise achievement of the Aichi Targets as well as national NBSAP goals. The reasons for these actions is that the time required to reach an ABS agreement is likely to damage research significantly; such delays for access to physical specimens are already problematic, and there is no reason to expect that the very much higher rate of use of NSD will be handled any more expeditiously. To put this into context, every weekday, well over 38 million requests are made to EMBL-EBI websites⁷.

3. Benefit-sharing arrangements from non-commercial use of digital sequence information on genetic resources.

In our submission of 2017 we discussed the value of “DSI” to the implementation of the CBD, and would withdraw nothing from that statement⁸, which was based almost entirely on the concept of

⁷ <https://www.ebi.ac.uk/about/our-impact>

⁸ <https://www.cbd.int/abs/DSI-views/CETAF-DSI.pdf>



NSD discussed here. Implementation of the CBD is of course on a national basis, and thus countries implementing the CBD are making use of benefits developed through the use and generation of NSD.

CETAF members agree bilateral benefit-sharing arrangements as a normal part of Mutually Agreed Terms when accessing genetic resources. Insofar as the benefits are data or information (as opposed to capacity-building, for example), practice is to both share them bilaterally and to publish them so they are available globally. While we have many examples of bilateral benefit-sharing in the context of tangible genetic resources, we have no experience of benefit-sharing on a bilateral contractual basis exclusively for NSD.

Although contract-based bilateral benefit-sharing for NSD is not in our experience, global benefit sharing from analysis of NSD is a common experience and part of daily work. As noted above, the prevailing model of non-commercial scientific work is of open sharing of data and outputs. As a concomitant to publication of research results researchers are required to make NSD from their own databases publicly available once research is published, generally through the large public databases of the INSDC. Thus benefit sharing when these benefits are new genetic sequences is done on a global (multilateral) basis rather than a bilateral basis. This methodology is far more efficient and valuable to all users, since it allows access to sequences relating to species outside national borders, important for identification of invasive species, for example⁹.

The global resource of NSD provided by the INSDC is in constant use worldwide; although the majority of users come from North America, Europe and China the data made available are used by every Party to the CBD¹⁰ (see <https://www.ebi.ac.uk/about/our-impact> for a real-time visualisation of use of EMBL databases). While researchers can (and sometimes do) send direct links to uploaded NSD to provider countries, there is no evidence that this makes any difference to availability and use of the data developed and shared.

NSD are in use globally, but there are still capacity building needs to increase Parties' ability to realise the benefits and exploit these data. The SCBD has supported training in DNA barcoding, which includes making use of the NSD in the BOLD system. INSDC members also offer training and a range of training materials. CETAF member organisations are also active in capacity building. This may take the form of training as a part of research, for example training students while working in labs in providing countries, joint research involving generation and analysis of NSD, in-house training at bachelor's, master's and PhD levels, and informally through professional contact. Many CETAF members also run DNA labs as a part of their infrastructure, and make these available to visitors and colleagues from developing countries, effectively increasing the capacity of those countries.

⁹ Lack of such sequences has been identified as a problem for invasive Alien Species detection – see Lyal & Miller, 2018, Capacity of United States federal government and its partners to rapidly and accurately report the identity (taxonomy) of non-native organisms intercepted in early detection programs. 22pp. https://www.doi.gov/sites/doi.gov/files/uploads/lyal_federal_capacity_taxonomy_contractorreport_22october2018.pdf

¹⁰ Leibnitz Association, 2018, The DSI debate: a primer on the science and infrastructure behind DSI. Discussion paper.



Summary

The term “Digital Sequence Information” is ambiguous and “DSI” is increasingly used as a convenient acronym without consideration of what it encompasses. We suggest that a replacement term be used in discussion and negotiation – ‘Nucleotide Sequence Data’ (NSD). This is the order in which nucleotides (Adenine, Thymine or Uracil, Guanine, and Cytosine) occur in a strand of DNA or RNA. The definition excludes ‘information’ which is developed through analysis of the data and which might be under the Intellectual Property Rights of the researcher. It also excludes ‘Digital’ to avoid restriction to a single storage medium.

Research increasingly involves generation of new NSD and downloaded NSD from public databases. While we have no examples of bilateral benefit-sharing on a contractual basis for use of NSD, the prevailing model of scientific publication of research results and the underlying data means that these results, and NSD, are available globally, the NSD being open access. We have seen evidence that users in all countries are accessing these data. This global availability of information to assist countries in implementing the Convention on Biological Diversity has been called for in a number of COP decisions and under Aichi Target 19.

We are aware that countries do not all have sufficient capacity to make full use of NSD. We regularly engage in capacity building through training and joint research, and see this as a continuing activity.

We are concerned that the development of restrictions on use of NSD will damage biodiversity research. Biodiversity loss is alarming¹¹, and further restrictions to identify and understand Biodiversity will generate massive drawbacks for the people’s well-being and all life on Earth and consequently endanger research in the conservation and sustainable use of Biodiversity.

CETAF ABS Core Group, 6th May 2019

¹¹ <https://www.ipbes.net/news/media-release-biodiversity-nature%E2%80%99s-contributions-continue-%C2%A0dangerous-decline-scientists-warn>