

Submission of views and information on benefit sharing arrangements from commercial and non-commercial use of digital sequence information on genetic resources

A submission from:

The Natural History Museum, Cromwell Road, London SW1 5BD, UK
Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AE, UK
Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh, EH3 5LR

In Response to notification 2019-012 on 5 February 2019:

The Conference of the Parties to the Convention on Biological Diversity, at its 14th meeting, in decision 14/20 requests Parties, other Governments, indigenous peoples and local communities, relevant stakeholders and organizations to submit their views and information on Digital Sequencing Information¹:

- (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.

Executive Summary

We recommend that term ‘Nucleotide Sequence Data’ (NSD) is used instead of “DSI”, and that this concept be taken as the basis for discussion. NSD is the DNA (or RNA) sequence - the arrangement of nucleotides² in a molecule of DNA or RNA. Our reasons are:

- i. A genetic resource when accessed is not accompanied by information about properties or other products of research. The concept of “DSI” closest to an accessed tangible genetic resource is the nucleotide sequence, without its properties or other information developed through research. Neither nucleotide sequence data nor a tangible genetic resource have intrinsic information but can only be considered as providing data;
- ii. Use of “data” avoids use of the term “information”, since information, as the result of research on a genetic resource, might attract intellectual property rights which would not *a priori* fall under providing country sovereign rights;
- iii. Inclusion of the term ‘Digital’ might be taken to limit the applicability of any discussion to content available digitally (i.e. in computer databases or online) only;
- iv. The term ‘Digital Sequence Information’ is not used in the taxonomic, systematic and conservation communities and not understood

Non-commercial scientific research, including that using NSD, increasingly uses an open access model. Funders and publishers demand that data are published and made openly available, and the public

¹ <https://www.cbd.int/doc/decisions/cop-14/cop-14-dec-20-en.pdf>

² Nucleotides are subunits that are polymerized (connected to form a linear chain) to make nucleic acids (DNA and RNA). The four types of nucleotides in DNA and RNA are Adenine, Thymine, Guanine, and Cytosine, which are usually abbreviated to A, T, G and C. In RNA Thymine is replaced by Uracil (U). The order in which these nucleotides occur in a strand of DNA is the DNA (genetic) sequence.

databases that hold the overwhelming majority of sequence data have an open access policy requirement. This model supports scientific research and environmental management globally; users of these databases are in every Party to the CBD. The open access model supports scientists in all countries, noting that research results become more reliable with increasing amount of data on which they are based. Open release of NSD for use by the global scientific community is clearly an example of benefit-sharing.

We recognise that a lack of capacity in many countries hinders greater use of NSD. Non commercial biodiversity research organisations are engaged in capacity building to meet this problem through training at graduate, postgraduate and professional levels, and by supporting infrastructural growth in provider countries. This is considered good practice within the sector.

Benefits derived from NSD use are mostly applicable on a multilateral basis; many projects make use of NSD generated from multiple geographic sources and publish on taxa that occur in more than one country. Unrestricted publication of data and literature is the most effective model for enabling this.

Our view is that the value of openly publishing NSD and thus making non-monetary benefits derived through their use³ available to users in all countries outweigh the risks of these data being used for commercial purposes without monetary benefits being shared with countries from which the original genetic resources were accessed. However, more research is needed to identify cost-effective solutions that mitigate the risks without increasing the risk of loss of non-monetary benefits.

Increasingly countries are including some form of “DSI” in their domestic ABS legislation. The non-commercial sector ultimately needs clarity on what is and is not covered by their usages. This issue cannot be ignored, since some national legislation and permitting procedures are beginning to restrict access to and use of genetic resources through these bilateral agreements, and consequently reducing the development of non-monetary benefits that can be shared.

Preamble

The three institutions submitting this document are committed to supporting implementation of the CBD and to benefit-sharing in our day-to-day operations. Our operations are primarily non-commercial, and our utilisation of genetic resources does not generate monetary benefits. Our benefit-sharing is consequently predominantly in terms of non-monetary benefits. For DSI as discussed below we believe that the most effective basis for benefit-sharing is on a global basis in the manner required by Aichi Target 19. This is already a model that is operating, as discussed below.

A summary of our previous submission on “Digital Sequence Information” and its importance to the CBD, made in response to notification 2017-37 issued on 25 April 2017⁴, is provided below:

- The generation of “DSI” is growing rapidly in terms of the number of individual organisms and species being sequenced and the depth of genomic coverage obtained per sample.
- Three global databases, known as the International Nucleotide Sequence Database Collaboration (INSDC), exchange information and currently mediate data on more than 200 million sequences; one of them (EMBL-European Bioinformatics Institute (EMBL-EBI), is subject to more than 100 million searches a year.
- Expecting the large-scale open-access international databases to regulate the use of ‘DSI’ from them is impractical and inappropriate: (i) the publisher cannot be responsible for compliance with any conditions by its users; (ii) the INSDC databases are required by their core policies (approved by national / regional authorities) to avoid both barriers to downloading the sequence data and applying conditions on their use.

³ See the 2017 submission from NHM, RBG Kew and RBG Edinburgh for a discussion of these benefits

⁴ <https://www.cbd.int/abs/DSI-views/NHM%20RBGK%20RBGE-DSI.pdf>

- Publicly-available ‘DSI’ provides a vital benefit for biodiversity conservation and sustainable management of natural resources and supports the implementation of the Convention. The INSDC databases form *de facto* a part of the Global Taxonomic Information System called for in COP decisions, and the sharing of knowledge and the science base required for the achievement of Aichi Biodiversity Target 19.
- No country holds sequence data for all of its biota and species likely to be intercepted by quarantine as Alien Species, pests etc. This would still be the case if sequence data were treated as a bilateral benefit between researchers and provider country. The only way in which Parties can obtain sequence data for supporting implementation of the Convention is through the freely-available global databases.
- Our unequivocal view is that sharing “DSI” without hindrance is overwhelmingly beneficial to Parties to the Convention, assisting them to understand and manage their biodiversity though, *inter alia*, establishing identities and names to taxa, understanding genetic structure of populations for effective conservation management, assessing the diversity of organisms present in a given site, detection of protected or invasive species, identification of species in trade, understanding migrations, effects of harvest rates on genetic diversity, understanding changes in genetic diversity over time, phylogenetic diversity across landscapes to assist large scale conservation planning, pollinator conservation.
- Any modification of the current model of use of “DSI” would risk limiting the non-monetary benefits indicated above, and consequently the implementation of the Convention. The financial equivalence of these benefits has not been assessed, but before any action is taken it would be helpful to make this calculation and compare it (plus the implementation costs) to the revenues that might be generated by alternative models. Monetary benefits may also not be realised, as legal uncertainty and audit complexity would likely discourage usage of biological materials.

We argue in this current document that “DSI” be replaced by the more precise term “Nucleotide Sequence Data” (NSD). Our previous comments using the term “DSI” apply to NSD.

On the concept, including relevant terminology and scope, of digital sequence information

The term ‘Digital Sequence Information’ is not used by taxonomic, systematic and conservation research and we find it unhelpful. There is no agreed meaning, and we are concerned that it, and similarly imprecise terminology, is beginning to be used in bilateral contracts (permits, ABS contracts etc) and national legislation, consequently introducing legal uncertainty. This is causing reluctance by scientists to carry out research in countries with such legislation. We explore below what the term means to us and propose an explicit meaning and possible alternative terminologies.

A tool now vital for taxonomy, systematics and conservation globally is the analysis of the arrangement of nucleotides on a strand of naturally occurring DNA or RNA – the ‘Nucleotide sequence’⁵. These are ‘Nucleotide Sequence Data’ (NSD). NSD include non-coding & coding sequences, regulatory sequences, conserved sequences, genes that encode specific traits, DNA without known function and ‘junk DNA’. Larger data elements would include the entire genome of an organism [or, indeed, of a clade (pangenome) or environmental sample (metagenome)]. NSD are aggregated from naturally occurring genetic resources generated as a part of research or downloaded from INSDC⁶ and other databases. Analyses of NSD are interpreted in research to develop understanding of biological diversity at genetic, species and ecosystem levels.

⁵ See footnote 2 above

⁶ International Nucleotide Sequence Database Consortium

We distinguish between ‘data’, which are observations of naturally occurring conditions lacking extrapolated meaning, and ‘information’, which emerges through cognitive (or other) processing and application of data - research. Since *information* is developed through processing or analysis of *data* it can potentially be covered by Intellectual Property Rights of the researcher and therefore *a priori* cannot come under sovereign rights of a country from which the data were sourced⁷. This is not to exclude the potential for benefit-sharing with that country⁸. Indeed, our practice is to share such information freely and we understand this as part of our benefit sharing activities. Consequently, we regard NSD as intrinsic factual data directly derived from the genetic resource. Any interpretation or annotation of these facts are here regarded as information, and as benefits that can be shared.

When a genetic resource is accessed *in situ* in a provider country, all that is acquired is an organism or set of organisms. These lack a known identity, and there is no attached information; any information has to be provided through research or pre-existing knowledge of the researcher. Discussion of “DSI” in the context of sovereign rights would logically be an analogue of what is accessed as a material entity, focusing on the genetic resource element of that entity. For this reason we propose that the concept of “DSI” be restricted to NSD.

Our definition of “DSI” is therefore the arrangement of nucleotides on a strand of naturally occurring DNA or RNA, and whenever we refer to “DSI” this is what we intend. We note that this clear concept is in line with that of ‘Genetic Sequence Data’ (GSD) proposed by some Parties to the CBD. We propose that either ‘GSD’ be adopted as a working term, or the more precise ‘Nucleotide Sequence Data’ (NSD). We specifically exclude ‘digital’ from our proposed terminology to avoid inappropriate restriction to a single condition of storage or transmission medium of aggregated data from genetic resources.

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

Bilateral benefit-sharing arrangements are general practice when research is carried out on Genetic Resources accessed with PIC and MAT under the ABS requirements of many countries. Where NSD are generated through utilisation of Genetic Resources under such a traditional bilateral agreement and used subsequently in the same research then bilateral benefit-sharing arrangements may be in place. However, we have no experience of bilateral benefit sharing as a result of using exclusively publicly-available NSD.

In addition to bilateral benefit-sharing, the results of non-commercial research in our sector, and the data that underly those results, are made public, increasingly on an open access basis. Thus scientific data, information and research results are made globally available. Scientific research has always operated on a multilateral benefit-sharing model (although not termed as such). No Party to the Convention has the capacity to generate or manage data or information on all of its biota. Instead all must rely on data and information generated and held elsewhere. If scientific data and information were treated solely in a bilateral benefit-sharing manner, countries would not benefit from information generated about non-endemic species, or from *ex situ* collections. Data sharing is particularly applicable to NSD which, once used for research (and sometimes before research is published), are made publicly available through open access databases. In our previous submission we discussed these databases and noted that their development and accessibility was *de facto* a contribution to the global taxonomic information system as called for in CBD COP IX/22. We also showed that it is a response in part to CBD Parties’ calls for availability of information in a number of other COP decisions, including COP IX/22, COP X/39, COP XI/29 and the Aichi Biodiversity Targets, particularly Target 19. We reiterate that scientific research and use of NSD analysis for environmental

⁷ The Annex to the Nagoya Protocol specifically lists joint ownership of IPR in its list of non-monetary benefits, acknowledging that this is a benefit developed by research.

⁸ See World Intellectual Property Organisation (WIPO), 2018, A guide to intellectual property issues in Access and Benefit-sharing Agreements. WIPO, Geneva.

management, is increasingly effective with the increased quantities of available data. Global open-access systems are the most effective way of achieving this.

Publishing NSD held privately within our institutions is a form of benefit sharing, as well as a requirement of modern science. The uses to which these data are put in support of countries' activities to conserve and sustainably use their biodiversity were discussed in our previous submission. That NSD and other related data are being used globally can be seen from the EMBL-EBI's live data map⁹ which shows live use of their databases; a recent analysis showed that users every country on Earth made use of the INSDC databases¹⁰.

We are aware that there are capacity needs in some developing countries that limit use of NSD, although we are also aware of researchers in many countries making effective use of these data. We view capacity building in making use of NSD as a part of benefit sharing. We engage in this in several ways:

- Our institutions all run training courses including international students. These courses develop skills in DNA sequencing and analysing nucleotide data. This raises capacity in the students' countries to make use of NSD.
- Scientists in our institutions all support and nurture PhD students, including from developing countries, and some of these at least are engaged in the degeneration and analysis of NSD.
- Our institutions play host to research partners and visitors who make use of our molecular laboratories to carry out sequencing.
- Our laboratories regularly carry out sequencing tasks at cost for external researchers, both of tissues they submit and of tissues held in our own collections.
- Many of our researchers are in research partnerships with researchers in other countries, including developing countries, and will generate NSD as a part of their joint research. In this way, where sequencing facilities are unavailable elsewhere, we can provide that capacity at a far lower cost than would otherwise be required.

While the publication of NSD developed in our research is, we submit, a benefit, and one supplemented by our capacity building, the research results we publish arising from the use of published NSD also provide benefits. We discussed the use of such publications in the implementation of the CBD in our previous submission. As noted, publications from our researchers are made available globally. Increasingly we are seeking to publish in open access journals so that payment is not an issue. Because we work in a global community of research papers tend to be shared with our peers in both developing and developed countries, ensuring that research results are seen rapidly by scientists most able to use them. Such sharing of results is not documented in any standard way, but is a regular part of our activities.

A direct way in which we share benefits of NSD is in our use of it for identification. Every year, our staff identify numerous organisms for bodies in developing countries, often though the use of online data for comparison. Such identifications might not be possible without this resource.

Risks

We have presented evidence that use of NSD contributes to knowledge of biodiversity, both academic and applied in the context of the CBD. We are aware that there is concern among some countries that benefits (particularly monetary benefits) arising from the use of NSD are not being shared. However, we are unaware of any analysis of the actual value and development costs of commercial products generated following use of NSD. We are concerned that attempts to manage use of NSD will have a

⁹ <https://www.ebi.ac.uk/web/livemap/>

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

Implications of the use of DSI – NHMUK, RBG Kew, RBG Edinburgh

negative impact on non-commercial use and consequently on biodiversity research and environmental management.

Restrictions applied in some access permits, for example prohibiting the dissemination of sequence data to third parties, or extracting sequence data being only allowed by the GR collector are already having a negative impact. Researchers are unable to carry out their research effectively, and restriction on publishing means that biota from the country are excluded from the research. In addition to these constraints, some countries are explicitly stating their sovereign rights over “DSI” and bringing it into their domestic ABS legislation. As yet there is no clarity on how such legislation can be applied in practice, on the meaning of “DSI” or the legal implications of such legislation in other countries. Without some progress in international agreements we see problems in carrying out research, and sharing benefits as required by the CBD, increasing to the detriment not only to research but to biodiversity conservation and sustainable use.