

Comments of Third World Network on Digital Sequence Information

Reference: SCBD/NPU/DC/VN/KG/RKi/87804¹

1 June 2019

In response to the Executive Secretary's invitation pursuant to Decision 14/20, paragraph 9, Third World Network is pleased to submit the following to submit 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, and b) on benefit-sharing arrangements from commercial and non-commercial use of digital sequence information on genetic resources.

General

Technological advances in a cluster of linked sciences and technologies including gene synthesis, gene editing, cell culture, epigenetics, genomics-guided breeding, high throughput phenome/genome screening, and numerous other “-omics” applications are, indisputably, changing the ways in which genetic resources are used. Some of these technologies may prove useful for conservation and sustainable use (the first and second objectives), but it is with respect to access and benefit sharing (the third objective) that, for the foreseeable future, the heaviest impacts will be felt.

Without doubt, the technological reality for some years now is that an increasing number of genetic resources can be accessed as digital sequence information (DSI),² rather than as biological material, and that users can thereby avoid benefit sharing, both by synthesizing materials from sequences³ and by using the DSI itself for commercial purposes.⁴

For example, the pharmaceutical company Regeneron recently used the gene sequence of a Guinean Ebola virus collected in 2014 to create a treatment. Called REGN-EB3, the Ebola drug directly arises from use of Guinean genetic resources. It has received US \$400 million in research support and product orders from the US government. To make the drug, Regeneron downloaded the Guinean sequence information from GenBank and then synthesized portions of the C15 strain genome, using them to generate the monoclonal antibodies of which the patented therapy consists.

¹ Submitted by Edward Hammond, Advisor. Contact: Tel: +1 325 347 2829 | Email: eh@pricklyresearch.com

² Dormitzer et al. 2013. Synthetic Generation of Influenza Vaccine Viruses for Rapid Response to Pandemics. *Science Translational Medicine*, 15 May. DOI: 10.1126/scitranslmed.3006368

³ Hammond E 2019. Ebola: Company avoids benefit-sharing obligation by using sequences. Third World Network. May. URL: http://www.twn.my/title2/intellectual_property/info.service/2019/ip190504.htm

⁴ See, for example, the “CropOS” product of Benson Hill Biosystems, funded by Google Ventures, which relies on machine-based analysis of a variety of public and proprietary DSI. The system is sold to biotechnology and plant breeding institutions. URL: <https://bensonhillbio.com/design-better-crops-together-benson-hill-biosystems/crop-improvement-platform-cropos/>

The C15 gene sequence was placed in GenBank – which makes sequences available for “free”, no strings attached - by the Nocht Institute, a research center in Hamburg, Germany. While Nocht loaded the C15 sequence into a “free” database, for transfers of samples of Ebola viruses, Nocht uses a legally-binding material transfer agreement (MTA) noting Convention and Nagoya Protocol obligations, and requiring negotiation of a benefit sharing agreement with Guinea in the event of commercial use.

Thus, by downloading the C15 sequence from GenBank and then synthesizing it, rather than requesting a virus sample from Nocht, Regeneron did not sign the MTA requiring negotiation of a benefit sharing agreement, and the company is manufacturing the product for drug stockpiles in North America without benefit sharing to Africa and Guinea.⁵

Some regions and countries anticipated such developments and have access and benefit sharing (ABS) rules on information, and others are now taking policy and legal steps. More generally, however, typical “classic” ABS approaches, often centered on MTAs, are inadequate to ensure benefit sharing for DSI, and ABS approaches are in need of major updates.

The task of creating a system that ensures benefit sharing when genetic resources are utilized as digital sequence information (DSI), and synthesized from DSI, is an existential challenge to the Convention. If benefit sharing is not required of users of DSI, then effective implementation of the third objective of the Convention will prove impossible. Users, especially commercial users, will avoid benefit sharing, thereby undermining the Nagoya Protocol and the Convention itself.

Other international processes related to biodiversity and the Convention look to the CBD for guidance on DSI. These processes, on genetic resources for food and agriculture and pathogens, have important ramifications for biodiversity that is used in public health and agriculture. The progress of these processes may depend on coordination with the Convention’s DSI approach and signals the importance of the Convention and Protocol rapidly moving forward at the next COP/COP-MOP.

Moreover, speed is of the essence because of technology-related concerns. DSI, particularly genetic sequence data, is presently being accumulated at a rate faster than it can be fully leveraged. This is because the ability to sequence is maturing more rapidly than the many ways to use sequence data emerging from genomics and other sciences. As a result, for some commercial purposes, accumulating sequences is akin to depositing money in the bank, or filling a petrol tank – it can be confidently foreseen that even more future uses will emerge for a resource gathered now.

But once such data is sequestered in a private databases, or placed in the (alleged) “public domain” through irresponsible forms of “open access” that do not respect the rights of genetic resource providers, the ability of Parties and IPLCs to protect their sovereign and traditional rights is impaired. The longer it takes for the CBD to act on DSI, the greater the problem becomes.

⁵ See Hammond 2019, *ibid*, for more information.

Terminology and Scope

With respect to terminology, we note that it is agreed that “Digital Sequence Information” (DSI) is currently used as a placeholder term. Importantly, what the term is agreed to include, that is, its “scope”, is more significant than the words of the term itself. Adapting or even creating a term *de novo* and assigning a definition to it is within the normal scope of implementation of binding agreements, thus, terminology *per se* is not the trouble here. Rather, it is agreement on the breadth of the term.

Whatever term(s) is/are ultimately used, DSI should necessarily include DNA and RNA sequences in all their forms, including assembled and annotated genomes and partial sequences, as well as sequences of alternative forms such as cDNAs, codon optimized sequences, etc. DSI should also include amino acid sequences, SNPs, STR counts, and epigenetic and molecular characterization information (e.g. structures, DNA methylation, etc) and sequence associated metadata (e.g. ‘passport’ data, phenome-genome data, etc).

We note that a useful contribution to the discussion of the various elements of DSI in the context of agriculture - many of which can also be related to biodiversity more generally - can be found in the *Draft Exploratory Fact-Finding Scoping Study on “Digital Sequence Information” on Genetic Resources for Food and Agriculture* prepared for the Commission on Genetic Resources for Food and Agriculture in 2018.⁶

We concur with others who believe that the word “digital” is both unnecessary and potentially confusing. It should not be part of the term(s) that is/are ultimately used, due to developments such as quantum computing and the much older phenomenon of storing sequence and other relevant information on non-digital media, such as plain old paper.

We understand DSI as a neologism that originally combined the phrases “digital sequences” with “genetic sequence information”, that were used in the 2015 meeting of the CBD’s Synthetic Biology AHTEG. At that meeting and in other early discussions, use of the word “digital” typically was not focused on the storage format of information, but rather sought to capture the comparative speed and ease with which modern information technology allows sequences to be transmitted across borders, often via the “digital” Internet. DSI is then stored, used *in silico*, and/or subsequently synthesized, and these phenomena pose considerable challenges to methods currently used to implement access and benefit sharing obligations, both in agriculture and beyond.

Thus capturing the above idea – the speed at which large amounts of DSI may be transmitted globally - in the term ultimately used is consistent with how the term arose and would be preferable to the less descriptive “digital”.

⁶ Heinemann J and D Coray 2018. Draft Exploratory Fact-Finding Scoping Study on “Digital Sequence Information” on Genetic Resources for Food and Agriculture. Commission on Genetic Resources for Food and Agriculture, CGRFA/WG-AqGR-2/18/Inf.10. URL: <http://www.fao.org/fi/static-media/MeetingDocuments/AqGenRes/ITWG/2018/Inf10e.pdf>

Benefit-sharing arrangements from commercial and non-commercial use of DSI

Benefit-sharing arrangements for commercial and non-commercial use of DSI should reflect the same or similar benefit sharing obligations as those attached to biological materials. Benefit sharing obligations should apply to both use of DSI that remains *in silico*, for example, to guide and inform marker-assisted breeding in agriculture, as well as use of DSI to generate materials, for example, vaccine viruses and cell culture strains, or the replication of DSI in material by gene editing.

As is the case with MTAs, *a priori* exemptions from benefit sharing obligations for DSI for categories of users are impractical given the legal structures of non-profit research and intellectual property systems in many countries, particularly developed countries. Rather, benefit sharing for DSI should be structured such that obligations are triggered based on types of use, without regard for the identity of the user, be it a company, a non-profit, an academic, or other entity.

For example, it would be unwise to exempt public or private academic users from benefit sharing obligations when they use DSI because in many countries, as a legal condition of employment, and as a condition of receiving public funding, academics are required to patent, and their institutions legally obligated to financially exploit, any commercially-applicable discovery, even if that discovery is incidental. Moreover, others may derive commercial uses from DSI that is initially handled by academic entities.

The US Bayh-Dole Act is an example of a law creating the above type of requirements, and Bayh-Dole Act types of legal obligations and institutional policies permeate the developed world (and some developing countries). Where such laws are in place, put simply, it is against policy or even illegal for academics not to patent and commercially exploit academic inventions, even if they are made accidentally.

While aspects of policy and law vary from country to country, across the North, it can consistently be observed that academic users of DSI:

1. Must report all inventions to their employer, even if unintended;
2. Must convey rights to inventions (or proceeds from them) to the employer, either as a matter of law or terms of employment;
3. Receive payments as personal income from their employer in return for their compliance;
4. May not waive or negotiate the intellectual property interests and policies of their institution in an access and benefit sharing agreement.

Having said the above, while all users of DSI should be subject to benefit sharing requirements, not every use of DSI – indeed not the vast majority of uses outside of corporate confines – would necessarily trigger action in relation to those obligations, particularly if the use proves to be genuinely non-commercial. Thus, the benefit sharing

obligations incumbent on users of DSI can and should contain carefully crafted thresholds that ensure that genuinely non-commercial research will typically proceed without triggering financial benefit sharing obligations.

One way to cause such benefit sharing obligations to take effect is through the use of **data access and use agreements** that DSI users must agree to before accessing data, for example, DSI in internet-linked databases. Data access and use agreements offer the potential to permit DSI to remain publicly-accessible while protecting the interests of providing countries and IPLCs. Akin to the “terms and conditions” that accompany an airplane or train ticket, or utilities such as water and sewer services, data access and use agreements can be used to set forth benefit sharing obligations connected to publicly-accessible DSI.

Opponents of benefit sharing for use of DSI sometimes suggest, while seldom offering evidence, that data access and use agreements and similar approaches would inherently impede research. Review of online databases related to biodiversity, however, confirms that it is quite normal for databases to employ terms and conditions, and this includes databases specializing in widely diverging sets of biodiversity.

The Global Biodiversity Information Facility (GBIF)⁷ emphasizes “free” data sharing in a way that may be inappropriate for some DSI, however even this “free” database aggregator utilizes a “data use agreement” that “requires and implies agreeing” to a variety of stipulations. These include that the national law of Denmark (GBIF Secretariat host) governs the agreement, recognition of the “scope and application of Intellectual Property Rights and benefit sharing agreements as determined under relevant laws, regulations and international agreements,” that data providers may restrict access to data, and that “reasonable assurances” be made by data providers that prior informed consent has been obtained.

(The latter, and some other stipulations of the GBIF data use agreement, fall well short of protecting the benefit sharing interests of provider countries and IPLCs in DSI, however, the fact that such stipulations exist in the GBIF user agreement demonstrate that terms and conditions relevant to DSI are already in use.)

The Center for Australian National Biodiversity Research and the Australian National Botanic Gardens assert a Commonwealth copyright and other restrictions on data hosted at their websites,⁸ which include a number of biodiversity databases. These restrictions expressly prohibit commercial use, stating that, unless otherwise noted, “*Information and data on this server have **NOT** been placed in the public domain, but are provided for the personal non-commercial use of educators, students, scholars and the public,*” and that “*Any **commercial use** or publication of these resources without a licence from the custodians is strictly prohibited*” (emphasis in the original).

The “World Database of Key Biodiversity Areas”⁹ (“KBA”), managed by Birdlife International on behalf of a consortium including IUCN, WWF, and a number of other

⁷ <https://www.gbif.org/en/terms/data-user>

⁸ <http://www.anbg.gov.au/copyright.html>

⁹ <http://www.keybiodiversityareas.org/info/dataterms>

organizations employs a “terms and conditions of use” document that asserts copyright and ownership over the data, prohibits commercial use, prohibits reposting or redistribution, and prohibits commercial use of derivative works. This latter restriction has substantial implications with respect to intellectual property.

The Seabird Tracking Database,¹⁰ also managed by Birdlife, has terms of use that assert copyright and state that database users may not “*adapt, alter or create a derivative work from any BirdLife International content except for your own personal, non-commercial use.*” This and other terms “*shall take effect immediately on your first use*” of the website.

The Marine Metagenomics Portal,¹¹ based in Norway, takes a less restrictive approach than Birdlife, IUCN, WWF, and colleagues, utilizing a Creative Commons No Derivatives license. This allows copying and redistribution of the data, however, users may not redistribute the data if they “*remix, transform, or build upon*” it, and users “*may not apply legal terms or technological measures that legally restrict others.*” These restrictions have an effect of limiting any intellectual property claims. To enhance clarity, the terms of use add that “*Any genetic information is provided for research, educational and informational purposes only.*”

The Arabidopsis Information Resource (TAIR)¹² maintains a database of genetic and molecular biology data for the model higher plant *Arabidopsis thaliana*. This database has lengthy terms that include “*You may not utilize the Service if you are an employee (including part-time employees) or contractor of, or in any way acting on behalf of a for-profit entity.*” Access to portions of the data requires registration and a paid subscription. The terms may be amended at any time without notice to users.

In the public health field, a number of DSI databases linked to pathogens impose terms and conditions on users. In the field of public health, in contrast to claims that benefit sharing for DSI will impede data sharing, in several instances databases with terms and conditions designed to protect the interests of genetic sequence data providers have become lauded examples of sharing.

For instance, GISAID, the Global Initiative on Sharing All Influenza Data, employs a detailed user agreement.¹³ Among its provisions, it prohibits users from reposting data in other databases, states that data providers have not relinquished data ownership by making it available to GISAID users, and prohibits users from providing data to third parties. While GISAID predates the Nagoya Protocol and its user agreement does not require benefit sharing, the highly elaborated GISAID agreement, with its emphasis on protecting data provider interests, has proven very popular among influenza researchers.

¹⁰ <http://seabirdtracking.org/termsfuse>

¹¹ <https://mmp.sfb.uit.no/terms/>

¹² https://www.arabidopsis.org/doc/about/tair_terms_of_use/417

¹³ <https://www.gisaid.org/registration/terms-of-use/>

MeaNS, the World Health Organization's database of measles infection data and measles strain gene sequences, operated by Public Health England,¹⁴ utilizes terms and conditions that prohibit the downloading of sequence data: *"...as part of the terms and conditions of site usage, downloading sequences is not allowed. This is to ensure that users who submit sequences will not have sequences used by other individuals without permission."* Some DSI in MeaNS is available in other, less restrictive, databases, but other data are not. The MeaNS terms further require database users to be academic or non-profit researchers and state that users may not use the database for *"reproduction, adaptations, preparation of derivative works, or distribution of copies of any portion of MeaNS for any purpose."*

RubeNS, the World Health Organization's rubella database, also operated by Public Health England,¹⁵ utilizes similar restrictions.

Thus, for databases hosting DSI of a wide array of biodiversity, including the databases noted here, agreement to terms and conditions is frequently required before access to information is granted, and those terms and conditions frequently impose particulars on commercial use and govern DSI including gene sequences.

It is worthwhile to explore how making access to DSI contingent on acceptance of data access and use agreements, and how such agreements can protect the rights of providers, can be means by which to operationalize benefit sharing for DSI and thereby protect the Convention and advance implementation of its third objective and the Nagoya Protocol.

¹⁴ MeaNS Terms and Conditions: http://www.who-measles.org/Public/Web_Front/terms_conditions.php Also see the FAQ: http://www.who-measles.org/Public/Web_Front/faq.php

¹⁵ <http://www.who-rubella.org/tanc>