



# Digital Sequence Information (DSI) : What are we talking about and what is at stake in the current discussions?

**Nagoya Protocol on access & benefit sharing  
Pacific implementation update webinar  
*Tuesday 29 September 2020  
2 pm – 3pm, Apia Samoa.***

Olivier Rukundo , ABS Capacity Development Initiative

”

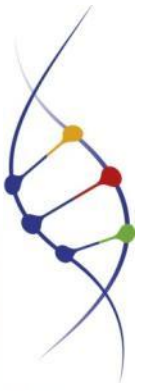
The ABS Initiative is funded by



and implemented by



# What is DSI?



- Technological advances are leading to the generation of biological data and information which is placed in databases. These databases are open access.
- Genetic sequence information data being produced and shared is what is being referred to as DSI under the CBD
- This data/ information is being increasingly used for research, R&D and commercial purposes
- With this, no need to access material in a physical manner.
- All this exchange of information is happening in the absence of benefit-sharing obligations : to date the major databases in which DSI is stored do not have terms and conditions on benefit sharing and broader access and utilization requirements.

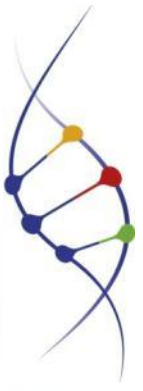


# What is DSI?

- No internationally agreed definition
- Emerged from discussions on synthetic biology
- Concept and scope are not clear
- More appropriate terminology is needed –placeholder
- Other terms in other fora (i.e. genetic sequence data)



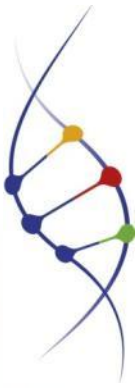
# Diverging views and positions on the issue



- Divergent positions on DSI and what it means
- On whether the definition of “genetic resources” and “utilization” of the NP include DSI
- On whether DSI should fall under the ABS regime of the CBD and the Nagoya Protocol
- On whether open access to DSI is a sufficient form of benefit-sharing



# Discussion on the issue happening in different fora



- UNCLOS -Biodiversity Beyond National Jurisdiction
  - Terms; use of marine genetic resources
- FAO –Commission on Genetic Resources for Food and Agriculture
  - Studies -understanding DSI use in different sectors
- FAO –International Treaty on Plant Genetic Resources for Food and Agriculture
  - One of the issue in the discussions on the Enhancement of the Multilateral System of the Treaty.

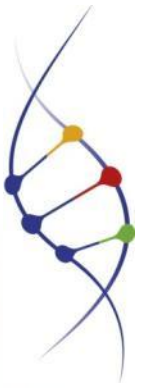


# Process at the CBD

- Decision 14/20 adopted at the 14th meeting of the Conference of the Parties to the Convention on Biological Diversity (COP14), established a science- and policy-based process to address digital sequence information on genetic resources (DSI) during the intersessional period leading up to COP15 in 2020.
- Process focused on the extension of an Ad Hoc Technical Expert Group (AHTEG) on DSI- initially established at COP 13



# Process at the CBD



- ❑ The AHTEG was mandated to, inter alia,
  - ❑ develop options for operational terms and their implications to provide conceptual clarity on DSI,
  - ❑ identify key areas for capacity-building concerning DSI, and
  - ❑ submit its outcomes for consideration by a meeting of the open-ended working group (OEWG) on the Post-2020 Global Biodiversity Framework.
  
- ❑ The OEWG's recommendations to COP15 are anticipated to address DSI in the context of the post-2020 global biodiversity framework.



# Process at the CBD: Main steps/ outcomes



## ❑ Synthesis of views to

- to clarify the concept, including relevant terminology and scope of DSI;
- if and how domestic measures on access and benefit-sharing consider DSI;
- benefit-sharing arrangements from commercial and non-commercial use of DSI; and
- capacity-building needs regarding the access, use, generation and analysis of DSI, in particular for the three objectives of the Convention.





# Process at the CBD: Main steps



## ❑ Commissioned studies

- Study 1 on DSI Concept & Scope: a peer reviewed science-based fact-finding study on the concept and scope of digital sequence information on genetic resources and how digital sequence information on genetic resources is currently used, building on the fact-finding and scoping study prepared for COP 14.
- Studies 2 & 3 on Databases & Traceability: a combined peer reviewed study on (a) ongoing developments in the field of traceability of digital information, including how traceability is addressed by databases, and how these could inform discussions on digital sequence information on genetic resources; and (b) public and, to the extent possible, private databases of digital sequence information on genetic resources, including the terms and conditions on which access is granted or controlled, the biological scope and the size of the databases, numbers of accessions and their origin, governing policies, and the providers and users of the digital sequence information on genetic resources and encourages the owners of private databases to provide the necessary information.



# Process at the CBD: Main steps



## ❑ Commissioned studies (ctnd)

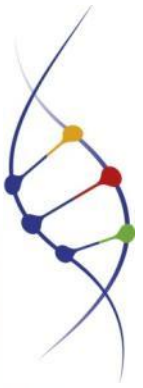
- *Study 4 on Domestic Measures*: a peer reviewed study on how domestic measures address benefit-sharing arising from commercial and non-commercial use of digital sequence information on genetic resources and address the use of digital sequence information on genetic resources for research and development, taking into account the submission of views received (see Synthesis above).

## ❑ AHTEG meeting

A meeting of the AHTEG on Digital Sequence Information on Genetic Resources took place ( virtually) in Montreal, Canada, from **17 to 20 March 2020**.



# AHTEG meeting: key outcomes



## ❑ AHTEG meeting ( Ctnnd)

- The outcomes of the meeting of the AHTEG are required to be submitted to the Open-ended Working Group on the Post-2020 Global Biodiversity Framework, established by the Conference of the Parties under its decision [14/34](#).
- The Working Group was requested to consider the outcomes and to make recommendations to the Conference of the Parties at its fifteenth meeting on how to address digital sequence information on genetic resources in the context of the post-2020 global biodiversity framework.



# AHTEG meeting: key outcomes



## □ AHTEG meeting ( Ctnnd)

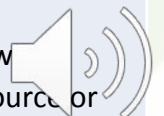
- **Distinction between data and information:** The experts noted that the latter could imply more processing than the former, and noting also that there is no clear boundary between the two.
- **Biochemical flow of information within a cell:** the experts noted the ability to infer nucleic acid sequences from protein sequences (although imperfectly, because different DNA sequences may code for the same protein) and the current inability to readily infer nucleic acid or protein sequences from metabolites and macromolecules.
- **Distinction between genetic and biochemical information:** Experts noted that associated information related to a genetic resource, such as traditional knowledge associated with genetic resources, and other information, variously described as contextual, associated, or subsidiary information



# Clarifying the scope of digital sequence information on genetic resources: 4 groups as a starting point



	Information related to a genetic resource			
	Genetic and biochemical information			Associated information
Group reference	Group 1	Group 2	Group 3	
High-level description of each group	DNA and RNA	Group 1 + proteins + epigenetic modifications	Group 2 + metabolites and other macromolecules	
Examples of granular subject matter	<ul style="list-style-type: none"> <li>· Nucleic acid sequence reads;</li> <li>· Associated data to nucleic acid reads;</li> <li>· Non-coding nucleic acid sequences;</li> <li>· Genetic mapping (for example, genotyping, microsatellite analysis, SNPs, etc.);</li> <li>· Structural annotation.</li> </ul>	<ul style="list-style-type: none"> <li>· Amino acid sequences;</li> <li>· Information on gene expression;</li> <li>· Functional annotation;</li> <li>· Epigenetic modifications (for example, methylation patterns and acetylation);</li> <li>· Molecular structures of proteins;</li> <li>· Molecular interaction networks.</li> </ul>	<ul style="list-style-type: none"> <li>· Information on the biochemical composition of a genetic resource;</li> <li>· Macromolecules (other than DNA, RNA and proteins);</li> <li>· Cellular metabolites (molecular structures).</li> </ul>	<ul style="list-style-type: none"> <li>· Traditional knowledge associated with genetic resources</li> <li>· Information associated with digital sequence information Groups 1, 2 and 3 (for example, biotic and abiotic factors in the environment or associated with the organism)</li> <li>· Other types of information associated with genetic resource or its utilization.</li> </ul>



# Scope of Digital Sequence Information



- **Scope of digital sequence information:** the AHTEG agreed that the first three groups proposed in Study 1 could be considered as digital sequence information, while associated information previously assigned (in the study) to Group 4, including traditional knowledge associated with genetic resources, is not digital sequence information



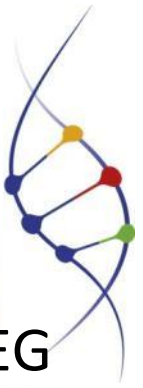
# Options for terminology to describe digital sequence information on genetic resources

Group reference	Group 1	Group 2	Group 3	Associated information
Category/term	<ul style="list-style-type: none"> <li>· Nucleotide sequence data (NSD);</li> <li>· Genomic sequence information;</li> <li>· Genomics information;</li> <li>· Nucleotide sequence information (NSI);</li> <li>· Genetic Resource Sequence Data (GRSD);</li> <li>· Digital sequence data (DSD);</li> <li>· Data on the genomic DNA (or RNA) of a sample genetic resource</li> </ul>	<ul style="list-style-type: none"> <li>· Genomic and proteomic sequence information;</li> <li>· Genomic and proteomic sequence information</li> <li>· Nucleotide sequence information (NSI);</li> <li>· Genetic information (GI);</li> <li>· Sequence data;</li> <li>· Nucleotide and amino acid sequence data (NASD);</li> <li>· Nucleotide and amino acid sequence and structural information (NASSI);</li> <li>· Nucleotide and amino acid sequence, structural and functional information (NASSFI);</li> <li>· Functional digital information of NSD;</li> <li>· Proteomic data;</li> <li>· Genomic and proteomic sequence information;</li> <li>· Data on the macromolecular composition of a sample genetic resource.</li> </ul>	<ul style="list-style-type: none"> <li>· Genomic, proteomic and metabolomic information;</li> <li>· Genetic and “omic” information;</li> <li>· Metabolomic data;</li> <li>· “Omic” information</li> <li>· Genomic, proteomic and metabolomic information;</li> <li>· Data on the biochemical and genetic composition of a sample genetic resource.</li> </ul>	<ul style="list-style-type: none"> <li>· Associated information ;</li> <li>· Contextual Information ;</li> <li>· Subsidiary Information</li> </ul>



Other terms were additionally discussed, including the following: digital sequence information, natural information, digital genetic resource information, digital genetic resource data and information, genetic resource data and information, genetic information, all data on a sample (genetic resource) and in silico

# Potential Implications arising from different groups



For each of the groups considered in the above table, the AHTEG discussed implications:

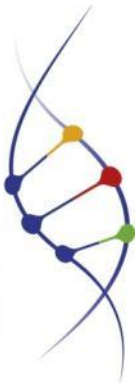
- (a) concerning traceability of different types of information;
- (b) concerning the use of digital sequence information and technologies enabled by digital sequence information in life sciences research and innovation processes;
- (c) of the International Nucleotide Sequence Database Collaboration (INSDC) on the open exchange and use of digital sequence information; and
- (d) concerning measures governing access, benefit-sharing and compliance.





## Potential implications of different groups concerning traceability and INSDC

- Experts noted potential ways to improve traceability, such as:
  - Enhancing the inclusion of relevant passport data (for example, by requiring the provider country field entries when uploading relevant records to the databases);
  - Including information regarding the genetic resource in databases;
  - Linking journal publications with genetic resources stored in *ex situ* collections.



# Potential Implications arising from different groups

## Potential implications of different groups concerning traceability and INSDC (ctnd)



- Some experts also considered:
  - Including disclosure of the provider country in patent applications;
  - Enhancing bioinformatic tools to support traceability, for example by direct comparison of sequences;
  - Explore the feasibility to link internationally recognized certificates of compliance (IRCCs) to genetic sequences uploaded in INSDC, including through interoperability.

Experts noted :

- **Traceability** could be more or less relevant depending on the approach to benefit-sharing followed, and the related monitoring and compliance requirements
- **In relation to Implications for INSDC:** it would be important to receive direct input from INSDC on this issue



## Potential Implications arising from different groups

### Potential implications of the different groups for technologies and/or sectors in the life sciences

- Experts noted that **discussions on implications for life-science sectors was of a preliminary nature** and that evaluating the implications of the different groups for the scope of digital sequence information would benefit from a deeper discussion
- Experts highlighted the **importance of having legal certainty** regarding usage of digital sequence information for all sectors therefore any approach to benefit-sharing should provide legal certainty, incentivize the use of digital sequence information and decrease unnecessary burden in monitoring, tracing, and tracking requirements.



# Potential Implications arising from different groups



Potential implications of the different groups or options concerning measures governing access, benefit-sharing and compliance

- **The multiplicity of different ABS national frameworks addressing digital sequence information on genetic resources poses challenges for users**, including those involved in basic non-commercial research, academic research and small and medium-sized enterprises.
- Discussion on potential implications for the different groups concerning measures governing access, benefit-sharing and compliance **was of a preliminary nature, and it was noted that this issue would benefit from further discussion**
- Implications of the different groups concerning measures governing access, benefit-sharing and compliance **would depend on the different approach to benefit-sharing that might be taken**





THANK YOU FOR YOUR KIND ATTENTION !

