"What is DSI and What is it Not? (Options and Views)

Margo A. Bagley
Asa Griggs Candler Professor of Law
Emory University School of Law

EMORY UNIVERSITY

Digital Sequence Information (DSI) and the three objectives of the CBD

1) Conservation of biological diversity;

2) The sustainable use of the components of biological diversity; and

3) The fair and equitable sharing of the benefits arising out of the utilization of genetic resources.

DSI relates to all three (can be helpful in 1 and 2) but perhaps the most controversy relates to 3, fair and equitable benefit-sharing

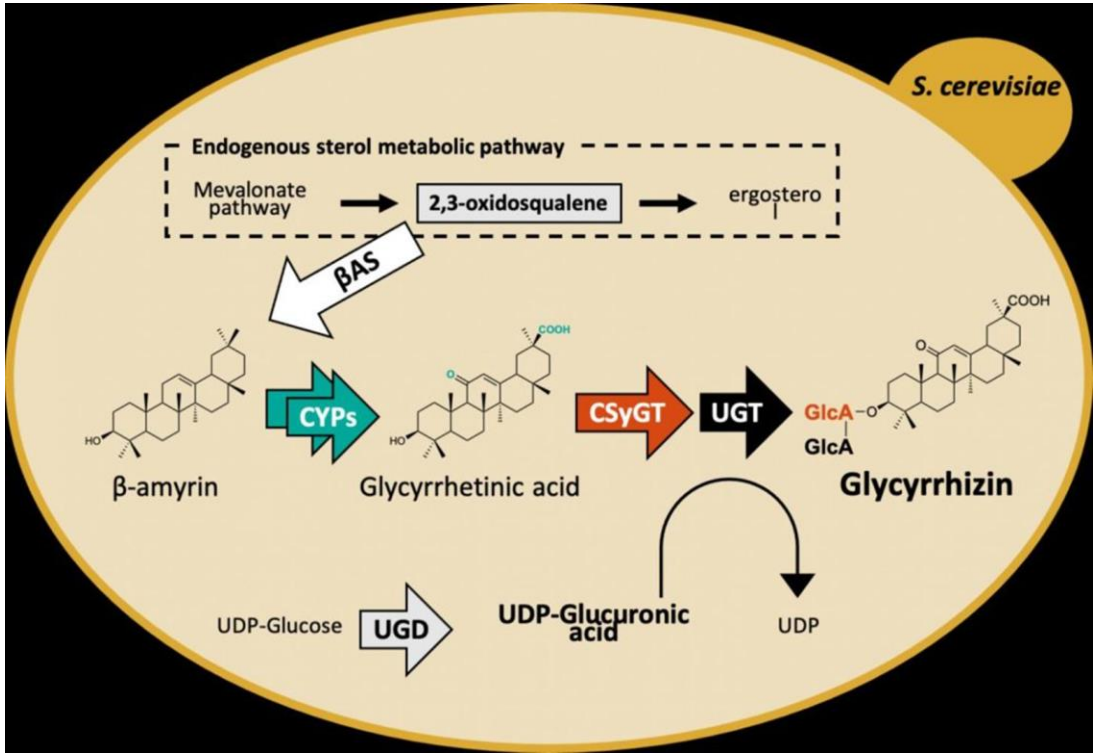
EMORY UNIVERSITY

Growing concern that “DSI” utilization will negatively affect Benefit-sharing by allowing users to bypass ABS obligations

Sequences obtained from **non-commercial** research may be made available in databases and used (without benefit sharing) in commercial projects (copying easy, cheap).

EMORY UNIVERSITY

Production of High Value Products without need for Tangible Material (can bypass benefit-sharing)



Engineered biosynthetic pathway for production of glycyrrhizin in yeast.

- "Enzyme Discovery Opens Door to New Methods for Producing Valuable Saponins"(Nov. 2020)
- improving commercial production of products which help reduce cholesterol levels, kill disease-causing bacteria, scavenge oxidative stress and inhibit tumor growth (found naturally in soybeans, peas, grains, herbs, etc.)
- "producing the chemicals in cell cultures would also reduce the need to deplete natural plant resources and help to meet sustainable development goals."

EMORY UNIVERSITY

Production of High Value Products without need for Tangible Material (can bypass benefit-sharing)

- "White Biotech Perfumery Will Continue to Grow" Nov. 2020

- Synthesizing high-value products through biocatalysis, metabolic engineering, synthetic biology, gene-editing and cloning, etc.
- Alternatives for sandalwood, patchouli, vetiver essential oil, and many other compounds already developed

What is Digital Sequence Information (DSI)? Many possibilities

2. The experts identified various types of information that may be relevant to the utilization of genetic resources, recognizing that the statements made further in the document might not apply equally to each of them. This included, among other things, the following:

(a) The nucleic acid sequence reads and the associated data;

(b) Information on the sequence assembly, its annotation and genetic mapping. This information may describe whole genomes, individual genes or fragments thereof, barcodes, organelle genomes or single nucleotide polymorphisms;

(c) Information on gene expression;

(d) Data on macromolecules and cellular metabolites;

(e) Information on ecological relationships, and abiotic factors of the environment;

(f) Function, such as behavioural data;

(g) Structure, including morphological data and phenotype;

(h) Information related to taxonomy;

(i) Modalities of use.

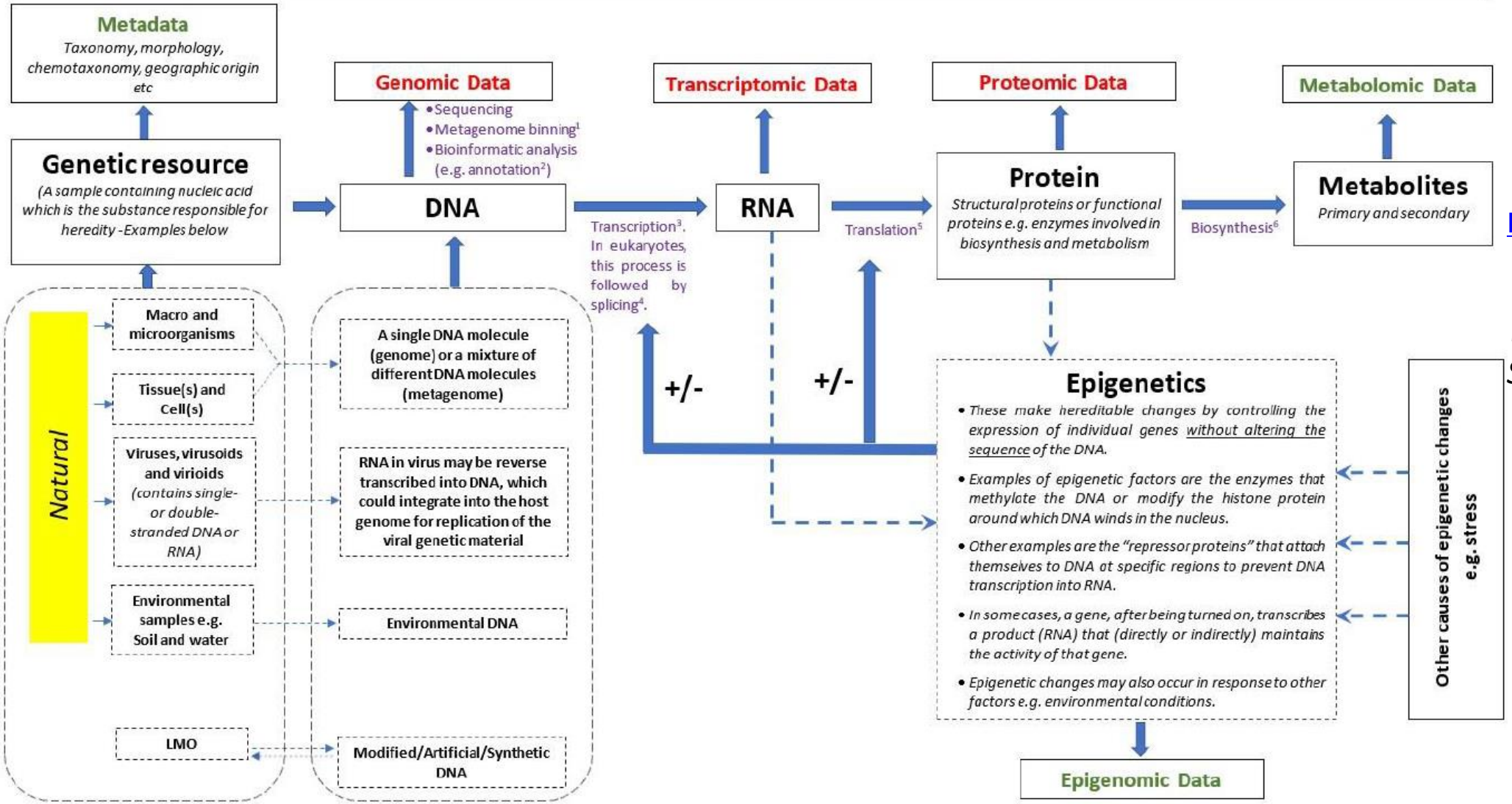
Report of the AHTEG on DSI, February 2018
https://www.cbd.int/doc/c/4f53/a660/20273cadac313787b058a7b6/dsi-ahteg-2018-01-04-en.pdf

What is DSI?

Decision 14/20 (COP 14) noted that DSI "may not be the most appropriate term" and is used as a placeholder until agreement is reached on an alternative term.

Much time and effort has been spent trying to ascertain the boundaries of the subject matter comprising DSI (issue will be under discussion at COP 15 in 2021).

Figure 1. Genetic resources and derivatives – flow of data and material.



DSI Study #1: Concept, Scope and Use (January 2020), *Wael Houssen, Rodrigo Sara, Marcel Jaspars*

Table 1. Clarifying the scope of digital sequence information on genetic resources

	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	• Nucleic acid sequence reads; • Associated data to nucleic acid reads; • Non-coding nucleic acid sequences; • Genetic mapping (for example, genotyping, microsatellite analysis, SNPs, etc.); • Structural annotation.	• Amino acid sequences; • Information on gene expression; • Functional annotation; • Epigenetic modifications (for example, methylation patterns and acetylation); • Molecular structures of proteins; • Molecular interaction networks.	• Information on the biochemical composition of a genetic resource; • Macromolecules (other than DNA, RNA and proteins); • Cellular metabolites (molecular structures).	• Traditional knowledge associated with genetic resources • 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) • Other types of information associated with a genetic resource or its utilization.

Report of the AHTEG on DSI, March 2020

"the degree of biological processing and the proximity to the underlying genetic resource provide a rationale to group information that may comprise DSI. . . . The proposed groups are cumulative. . . . Technological innovations might add to the granular subject matter." https://www.cbd.int/meetings/DSI-AHTEG-2020-01

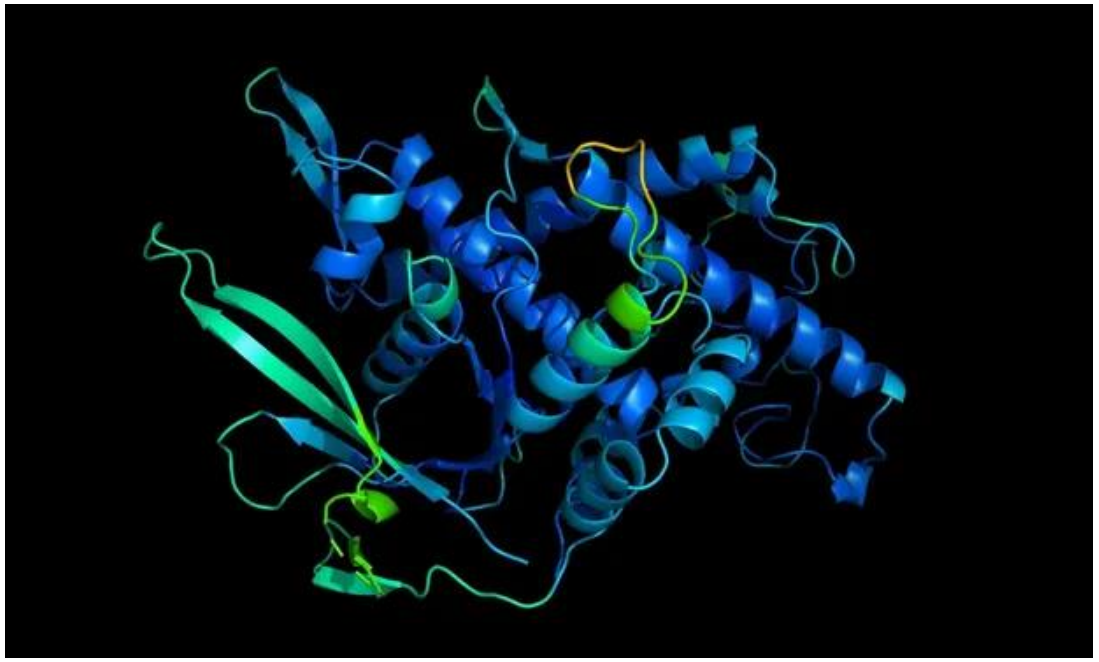
What DSI is not:

DSI AHTEG 2020 proposal: DSI is not associated information.

Table 2. Options for terminology to describe digital sequence information on genetic resources

Group reference	Group 1	Group 2	Group 3	Associated information
Category/term	• Nucleotide sequence data (NSD); • Genomic sequence information; • Genomics information; • Nucleotide sequence information (NSI); • Genetic Resource Sequence Data (GRSD); • Digital sequence data (DSD); • Data on the genomic DNA (or RNA) of a sample genetic resource	• Genomic and proteomic sequence information; • Genomic and proteomic sequence information • Nucleotide sequence information (NSI); • Genetic information (GI); • Sequence data; • Nucleotide and amino acid sequence data (NASD); • Nucleotide and amino acid sequence and structural information (NASSI); • Nucleotide and amino acid sequence, structural and functional information (NASSFI); • Functional digital information of NSD; • Proteomic data; • Genomic and proteomic sequence information; • Data on the macromolecular composition of a sample genetic resource.	• Genomic, proteomic and metabolomic information; • Genetic and "omic" information; • Metabolomic data; • "Omic" information • Genomic, proteomic and metabolomic information; • Data on the biochemical and genetic composition of a sample genetic resource.	• Associated information; • Contextual Information; • Subsidiary Information.
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*.				

Rapid Technological Advances Need Flexible Terminology



- "DeepMind AI cracks 50-year-old problem of protein folding" Nov. 2020
- "Scientists have identified more than 200m proteins but structures are known for only a fraction of them. Traditionally, the shapes are discovered through meticulous lab work that can take years."
- "When researchers know how a protein folds up, they can start to uncover what it does. . . . the breakthrough would help researchers tease apart the mechanisms that drive some diseases and pave the way for designer medicines, more nutritious crops and "green enzymes" that can break down plastic pollution."

Is DSI within the scope of the CBD/NP? Two Primary Approaches

DSI is a "Genetic Resource"

DSI results, directly or indirectly, from the utilization of "Genetic Resources"

Is DSI within the scope of the CBD/ NP?

- "**Genetic resources**" means genetic material of actual or potential value, and "genetic material" means <u>any material</u> of plant, animal, microbial or other origin <u>containing</u> functional units of heredity. (CBD Art. 2)
- "**Utilization of genetic resources**" means to conduct research and development on the <u>genetic and/or biochemical composition</u> of genetic resources, including through the application of biotechnology as defined in Article 2 of the Convention (NP Art. 2(c))
- "**Biotechnology**" as defined in Article 2 of the Convention means any technological application that uses biological systems, living organisms, or derivatives thereof, to make or modify products or processes for specific use." (NP Art. 2(d))

EMORY UNIVERSITY

"Fact-finding 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"

Margo Bagley, Elizabeth Karger, Manuel Ruiz Muller, Frederic Perron-Welch, and Siva Thambisetty

- **16 countries and one subnational jurisdiction[1] were identified as having domestic measures (legal, administrative and policy measures) in place addressing DSI, variety of approaches, including:**
  - **Some countries interpret genetic resources to include DSI and require PIC/MAT (possible access restrictions)[2]**
  - **Some countries impose benefit-sharing obligations on the use of DSI, but not access requirements[3]**

DSI Study #4

1. Bhutan, Bolivia, Brazil, China, Colombia, Costa Rica, India, Kenya, Malawi, Malaysia, Mozambique, Namibia, Panama, Peru, South Africa, Uganda, and Queensland, Australia

2. Bhutan, Malaysia, Peru, Bolivia, China, Colombia, Kenya, Mozambique, Oman, Peru, Uganda (not all actively imposing requirements).

3. Brazil, India, Malawi, South Africa

Is DSI within the scope of the CBD/NP? Implications of the Two Approaches

DSI is a Genetic Resource

- DSI subject to PIC and MAT so possible access limitations (NP Art. 6)
- Could hamper scientific developments (bilateral negotiations not conducive the speed of science, the ways DSI is used, the difficulties in detecting/tracing uses of DSI) (how identify benefits from undetectable uses)
- Flat fee access/licensing approaches may be possible (AHTEG 2020 report)

DSI results from the utilization of Genetic Resources

- Avoids access limitations
- Benefit-sharing required (still a challenge if track and trace necessary) (NP Art. 3)
- Benefit-sharing could be accomplished via open access approaches, e.g.:
  - Triggered by utilization or commercialization
  - A global multilateral benefit-sharing mechanism (perhaps under Article 10 of the Nagoya Protocol) which can include monetary and non-monetary benefits (AHTEG 2020 report)

Difference between DSI being a Genetic Resource and Resulting from Utilization of a Genetic Resource

- **Parties can impose access restrictions on genetic resources**: **"[A]ccess to genetic resources** for their utilization **shall be subject to the prior informed consent** of the Party providing such resources that is the country of origin of such resources or a Party that has acquired the genetic resources in accordance with the Convention, unless otherwise determined by that Party." Nagoya Protocol Art. 6
- **But utilization need only require benefit-sharing**: "This Protocol shall apply to genetic resources within the scope of Article 15 of the Convention and to **the benefits arising from the utilization of such resources**." Nagoya Protocol Art. 3

Conclusions

- Several possibilities for scope of DSI and terminology identified by the 2020 AHTEG; important for terminology to be flexible enough to accommodate rapid technological developments
- Additional issues include classification of DSI as a genetic resource or as resulting from the utilization of genetic resources and implications for access and (possible non-bilateral) benefit-sharing
- Current panoply of different domestic approaches to DSI scope and classification suggest need for resolution at the international level.

Questions?