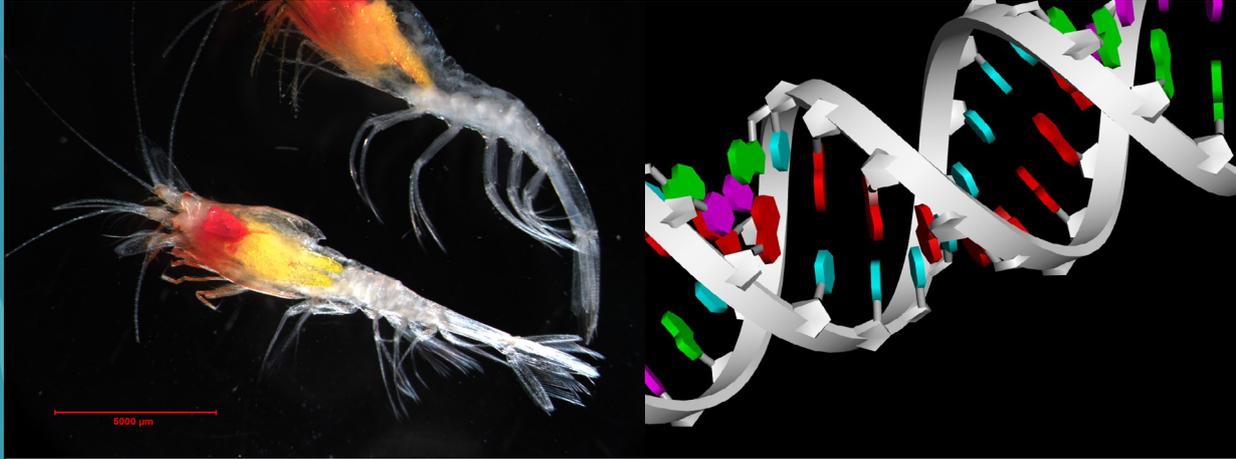


Digital Sequence Information – Clarifying Concepts



*Fig 1 The only pelagic shrimp known to inhabit the Canada Basin – Beaufort Sea, Alaska.
Photo credit: Hidden Ocean*

*Fig 2 DNA (computer generated).
Photo credit: Biovia Discovery Studio Visualiser 2019*

Despite “Digital sequence information (DSI)” forming a key part of discussions relating to genetic resources in processes under the Convention on Biological Diversity and the development of a new treaty for Biodiversity Beyond National Jurisdiction - there is no clear definition or common understanding of the term. Information demands a different benefit-sharing approach to materials. It is necessary to clarify the possible scope of DSI in order to inform development of benefit-sharing measures in the context of the conservation and sustainable use of marine biodiversity. Four possible subject-matter groupings) could be considered in developing a definition of DSI and ensuring scientifically meaningful terminology in international negotiations.

What is Digital Sequence Information?

Data, information and knowledge about genes represent an important element of benefit-sharing of genetic resources. “Digital sequence information” (DSI) on genetic resources is a ‘placeholder’ term used in discussions under the 1992 Convention on Biological Diversity. However, no consensus on a replacement or precise definition exists to date. In the development of a new international legally binding instrument for the conservation and sustainable use of marine biological diversity of areas beyond national jurisdiction, under the UN Convention on the Law of the Sea (BBNJ agreement), multiple terms have been suggested to replace DSI.¹ It is necessary to understand what ‘DSI’ could constitute.²

¹ *Digital sequence data, Digital sequence data and information, Genetic sequence data, Genetic sequence data and information, Digital genetic sequence data and information, Digital sequence information of genetic resources, Marine genetic resources in silico and [a combination of the above], Marine genetic resources in silico.*

² *This policy brief provides a summary of the more detailed study prepared for the CBD secretariat. https://www.cbd.int/a bs/DSI-peer/Study1_concept_scope.pdf*

Interpretations of the scope of DSI range from including DNA, RNA, and Protein sequences alone - to also including many other items only remotely connected with the original genetic resource (Figure 3). Discussions are ongoing to develop a more meaningful, well-defined term (<https://www.cbd.int/dsi-gr/>). A simplified flow of the potential scope of DSI, and techniques used to obtain this, derived from genetic resources is shown in Figure 3.

Why Should Materials and Information Be Treated Differently?

The problem with most attempts at applying access and benefit-sharing to information is that they have tried to replicate systems used for physical materials (e.g. samples of actual genetic resources) without recognising the fundamental differences between materials and information. Whereas physical materials are limited in quantity, using information does not stop anyone else from using the same information, and multiple users can access the same information at the same time. Often obtaining additional material may be expensive or impossible, whereas once the first copy of the information is made, the cost of making additional copies is marginal. It is easy to deny others access to materials, whereas once information is known, it is hard to exclude others from using it. These differences suggest that any access and/or benefit-sharing measures applied to information (i.e. DSI) should be different from those measures put in place for physical samples of genetic resources.

How Do Data and Information Differ?

A challenge faced in defining DSI, is in deciding what counts as data and the circumstances in which data is embedded with value and transformed into information. This distinction can be difficult to apply in practice; however, data is essentially a means of communicating and facilitating exchanges about the material world. Data describes inherent characteristics of material artefacts as distinguished from research outputs or other value-adding steps that generate knowledge such that its dissemination constitutes the sharing of information.

In the context of a genetic resource, the question arises as to whether 'DSI' should be confined to representational data (such as a DNA sequence 'GTACCTGA ...', methylation patterns, and so on) and if not, to what extent it should include processing activities performed with that data to generate information by data producers, curators and users. A key challenge is whether DSI, is limited to DNA and RNA sequences or whether it also captures the amino acid sequences of proteins and/or information generated by cognitive processes applied to such data. Different amounts of work are needed to convert different types of data into information. For instance, converting DNA sequences into their corresponding protein sequences is possible using a computer algorithm, whereas determining the three-dimensional structure of that protein requires a great deal of experimental effort.

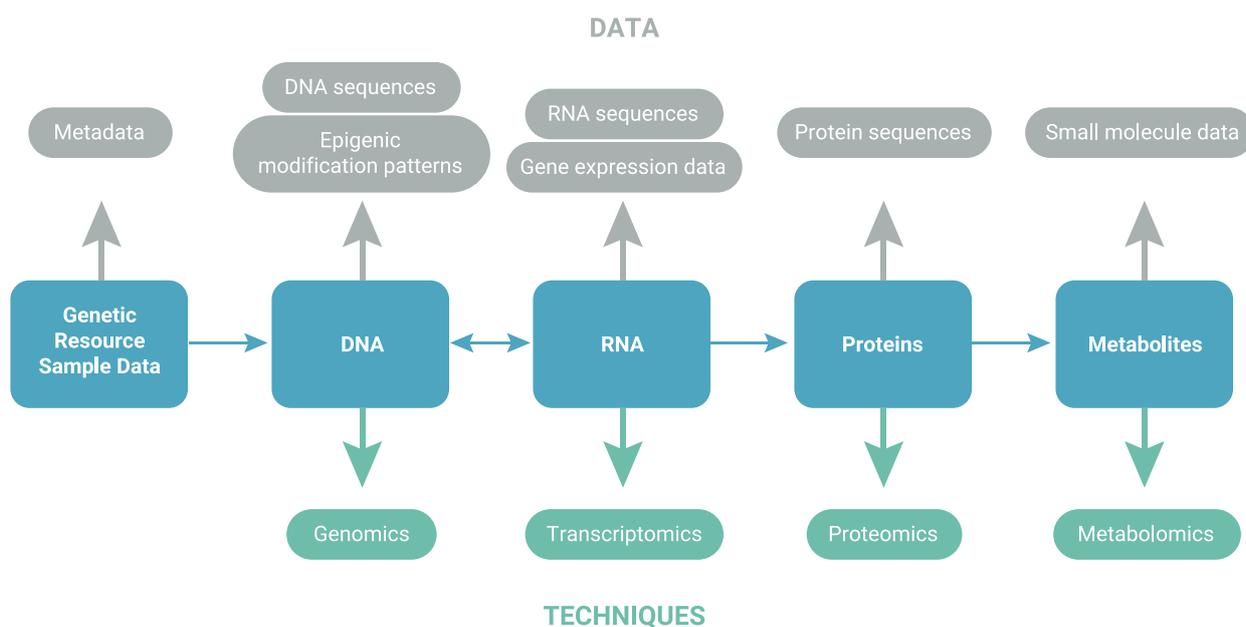


Fig 3 Digital sequence information on genetic resources and derivatives.

4 proposed groupings

Using the conceptual flow of data/information from a genetic resource (Figure 3), four possible interpretations of DSI can be identified (Figure 4), as follows:

- **Group 1** - Narrow: concerning DNA and RNA
- **Group 2** - Intermediate: concerning (DNA and RNA) + proteins
- **Group 3** - Intermediate: concerning (DNA, RNA and proteins) + metabolites and biochemical pathways
- **Group 4** - Broad: concerning (DNA, RNA, protein, metabolites) + traditional knowledge, ecological interactions, etc.

Scientifically, Groups 1-3 are all based on the molecular structure of macromolecules and small molecules, the information they carry, and information associated with their acquisition. Group 4 also includes information that is not related to molecular structure or information associated with their acquisition, with the weakest proximity to the underlying genetic resource, extending to behavioural data, information on ecological relationships and traditional knowledge. The groups differ in terms of biological processing carried out to convert data to information and (genetic?) proximity to the original genetic resource (Figure 5). Other, scientifically valid groupings could also be developed, and one such grouping is proposed by the International Nucleotide Sequence Data Collaboration which used 'Nucleotide Sequence Data' (DNA/RNA) and 'subsidiary information' (metadata and other data associated with the genetic resource and sequence data).

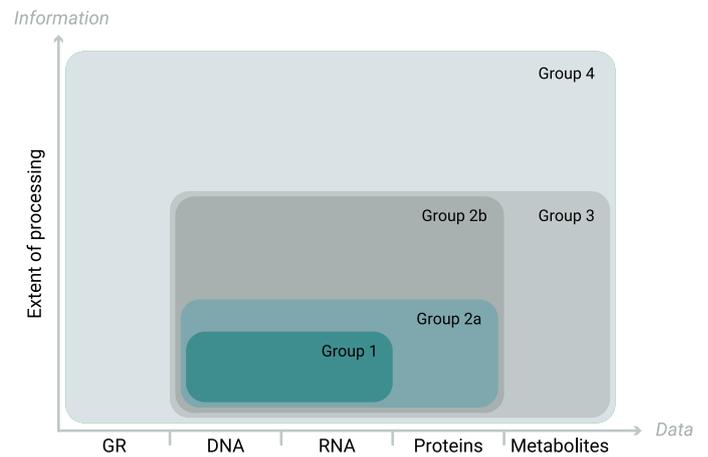


Fig 5 **Group 1** - Narrow: DNA and RNA;
Group 2a includes DNA/RNA sequence data including non-coding sequences, and information on the sequence assembly, including structural annotation and genetic mapping, as well as protein sequence data;
Group 2b is the same as group 2a in addition to which it includes functional annotation of genes, gene expression information, epigenetic data, and molecular structures of proteins;
Group 3 is the same as group 2b, but adds data on other macromolecules and metabolites, including their molecular structures;
Group 4 - Broad: DNA, RNA, protein, metabolites + traditional knowledge, ecological interactions, etc.

Looking ahead

Deliberations concerning the scope and concept of DSI can consider (i) the degree of biological processing and (ii) the proximity to the original genetic resource:

- How far along the flow from genetic resource onwards to DNA, RNA, protein sequences and metabolites 'DSI' can be considered to extend. Specifically: whether macromolecules (e.g. proteins, polysaccharides) are included under DSI and whether small molecules (metabolites) are included under DSI – *this can be resolved by utilizing the four groups proposed to clarify the scope of DSI subject matter, in which case all macromolecules (non DNA/RNA) and metabolites would be excluded under Group 1 or 2, whereas they would be included under Groups 3 or 4.*

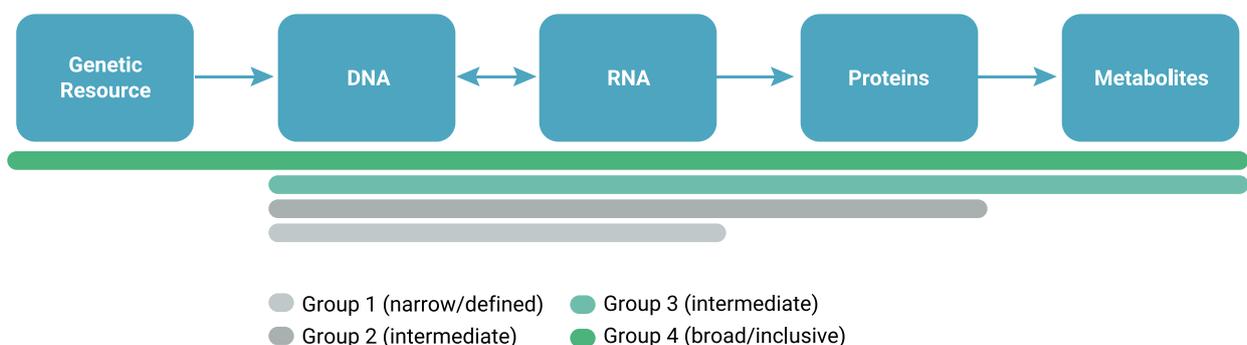


Fig 4 Proposed subject matter groupings for data/information potentially constituting DSI.

- ii. The distinction between data and information and how this is stored and processed, including the extent to which data has been processed before it can be considered information – *this can be resolved by utilizing the four groups proposed to clarify the scope of DSI subject matter as these have clear subject matter boundaries and so an approach, criteria or definition for distinguishing between data and information is not necessary.*

This Policy Brief is Based on:

“Digital Sequence Information on Genetic Resources: Concept, Scope and Current Use”. Authors:

Wael Houssem^{1,2}, Rodrigo Sara³, Marcel Jaspars¹

¹ Marine Biodiscovery Centre, Department of Chemistry, University of Aberdeen, Old Aberdeen AB24 3UE, Scotland, UK

² Institute of Medical Sciences, University of Aberdeen, Aberdeen AB25 2ZD, Scotland, UK

³ Consultant to the Secretariat of the Convention on Biological Diversity https://www.cbd.int/abs/DSI-peer/Study1_concept_scope.pdf

Also read: DOSI Policy Briefs:

“The Full Value of Marine Genetic Resources (MGR)”

“Accessing and Sharing Benefits from Marine Genetic Resources from Areas Beyond National Jurisdiction: Building on Best Practices in the Scientific Community”

ABOUT DOSI

The Deep-Ocean Stewardship Initiative seeks to integrate science, technology, policy, law and economics to advise on ecosystem-based management of resource use in the deep ocean and strategies to maintain the integrity of deep-ocean ecosystems within and beyond national jurisdiction.

For further information please contact:

dosi@soton.ac.uk

dosi-project.org

Additional input by Harriet Harden-Davies, Jane Collins, and Thomas Vanagt.

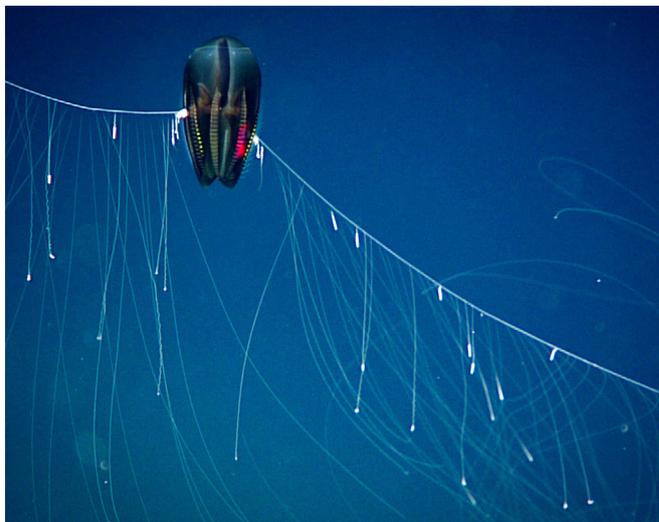


Fig 6 Dark ctenophore observed with its tentacles fully extended at approximately 1,460m deep, Gulf of Mexico. Photo credit: NOAA Office of Ocean Exploration and Research

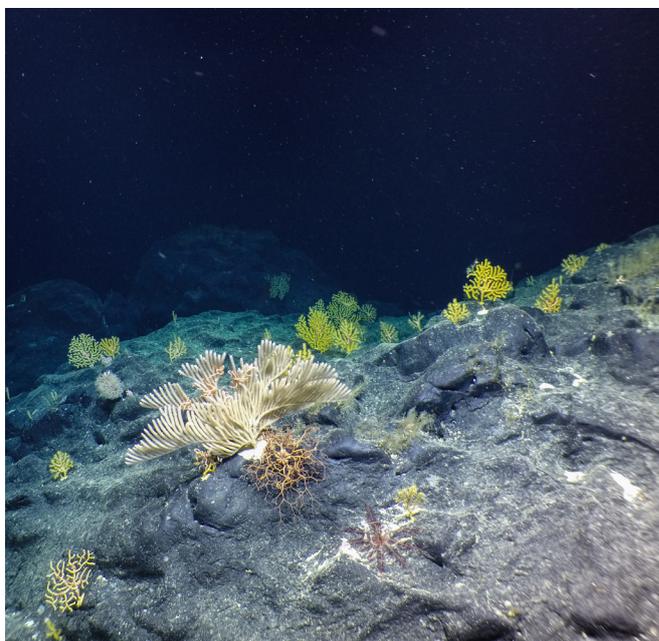


Fig 7 A mixed invertebrate assemblage on Jarvis Seamount #2 (central Pacific) at 1463m. Photo credit: Ocean Exploration Trust



Fig 8 Brisingid and corals. Courtesy of Malcolm Clark, NIWA, NZ