A Harmonized System for Benefit-Sharing from DSI

The Conundrum of DSI Benefit-Sharing

Sharing benefits which stem from the use of Digital Sequence Information (DSI) is now recognized as a critical topic by many governments and other stakeholders. It is currently the subject of discussions in four UN fora:

- At the 15th Conference of the Parties to the Convention on Biological Diversity (CBD), DSI was part of the Global Biodiversity Framework (GBF) package and Parties agreed to establish a multilateral mechanism and a global fund for benefit-sharing.
- DSI was also included in the recently adopted High Seas Treaty on biodiversity beyond national jurisdiction (BBNJ).
- Genetic sequence data under the Pathogen Access and Benefit-Sharing (PABS) system is also included in the Zero Draft text of the WHO Pandemic Preparedness CA+, currently under negotiation.
- Finally, DSI has been a long-standing, but unresolved, item under the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA).

The compartmentalized way in which DSI is being addressed in international policy fora reflects the approach by which national governments and the United Nations commonly structure responsibilities around differentiated sectors such as health, agriculture, and the environment. When applied to the development of a mechanism to share benefits from the use of DSI, this sectoral approach poses a number of potential practical challenges, since it does not reflect the way DSI is increasingly being made available and used in cross disciplinary, cross sectoral, research and development.

DSI data is stored in a variety of databases, and most researchers collect and use data that cut across different sectors, from agriculture, through human health to the conservation of wild species. For example, the International Nucleotide Sequence Database Collaboration (INSDC) is the largest repository of raw DSI data globally in which sequences generated from biological samples originating in all parts of the world are deposited. The INSDC does not separate out data from sovereign territories, marine habitats, plant species, or pathogens. Other databases can “specialize” in certain types of data, but vast amounts of data of different types are stored together in the larger databases. In fact, the most valuable databases for researchers are those that combine data, regardless of origin, because the data become useful when compared to other data. In isolation, a single segment of DSI has little value for researchers. Yet, with the parallel discussions under CBD, BBNJ, WHO, and ITPGRFA, multiple benefit-sharing mechanisms could apply to the same DSI data.

Why are multiple benefit-sharing mechanisms a problem?

A multiplicity of co-existing benefit sharing mechanisms for DSI could contribute to creating DSI silos and divergent rules on benefit-sharing:

- **DSI silos:** if harmonization of the different ABS systems is not achieved, the result could be subsets of DSI (e.g., plants, vertebrates, algae, pathogens) being housed in separate databases, or in the same databases, with different conditions for benefit sharing, affecting interoperability of databases. This compartmentalization would restrict, and even block, data from entering and freely flowing in and out of the global reference databases where most sequence data are currently stored and accessed. While this is not the intent, if different rules apply to different subsets of data, combining them into the larger databases could prove impractical and undesirable.
- **Divergent rules on DSI benefit-sharing:** if different policies require substantively different conditions to access and use DSI (e.g., clearing-house notifications during and after use of DSI, changes to database terms and conditions, or subscription/annual payments fees at certain trigger points), then users may struggle to understand what their legal obligations are and whether they have correctly complied with applicable rules. Regulators will similarly struggle to monitor user compliance when multiple paths towards compliance exist.

Both issues of silos and divergent rules have consequences for research. In concrete terms, these restrictions would affect:

- **Interoperability:** the usefulness of DSI depends on interoperability and integration of many types of data to facilitate meaningful analysis and conclusions. If DSI datasets are kept in separate compartments (i.e., in specific databases to ensure benefit-sharing according to specific provisions) or are subject to different rules within the same database, accessing and using DSI could be much more burdensome for researchers, limiting scientific enquiry. This is especially challenging if mixing benefit sharing models that link access to payment and models that do not.
- **Legal certainty:** Most research makes use of DSI that is in the remit of different international fora. If each of the international instruments governing DSI develops divergent benefit sharing mechanisms, it will be necessary to clearly define mutually exclusive scopes of application. Otherwise, a clear legal path for compliance will be elusive. As researchers frequently use thousands of DSI sequences, their datasets will most likely cut across several fora. Depending on the benefit sharing approaches adopted in those fora, it may not be clear which forum's benefit sharing rules, should apply. It is quite possible that researchers would be obliged to comply with two or more benefit-sharing mechanisms simultaneously, given the increasingly cross sectoral nature of much R&D. This will encourage risk-averse behavior and slow down research, which in turn would slow down benefit-sharing.
- **Compliance:** if users have to keep track of where DSI originated, it will require tracking and tracing subsets of DSI over the course of R&D processes. This monitoring of data is not always possible, but even when it is technically feasible, it would require complex compliance strategies for the hundreds of millions of DSI data. Such a system would incentivize "forum shopping," in which users preferentially use DSI that falls under a specific set of benefit sharing norms with simpler procedures at the expense of using the most appropriate DSI data. In some cases, users may decide to reduce the burden of compliance by avoiding use of DSI from certain fora if the system is too complex. The benefit sharing system must also take into account new developments which could render a system based on compartmentalized DSI obsolete, such as the use of artificial intelligence tools, which allow the "de novo" design of new synthetic sequences based on information contained in millions of sequences.

Research is the primary pathway through which benefits from DSI are realized. Benefits might come in a range of forms, such as generating new knowledge and innovations. If benefit-sharing systems undermine research or make it more difficult, in the long term they will ultimately reduce the benefits being created. If the system encourages opt-out or "forum shopping", it will also reduce the benefits being shared.

What is needed for a harmonized system to succeed?

A harmonized multilateral system for DSI access and benefit-sharing must be simple in order to ensure compliance and implementation, provide legal certainty to users, offer transparency on the benefits generated and cost less to enforce than the value of the benefits.

A critical window of opportunity now exists to ensure the various international policy processes involved in developing benefit-sharing mechanisms are compatible with scientific practices and DSI database structures, while maximizing benefits shared from the use of DSI.

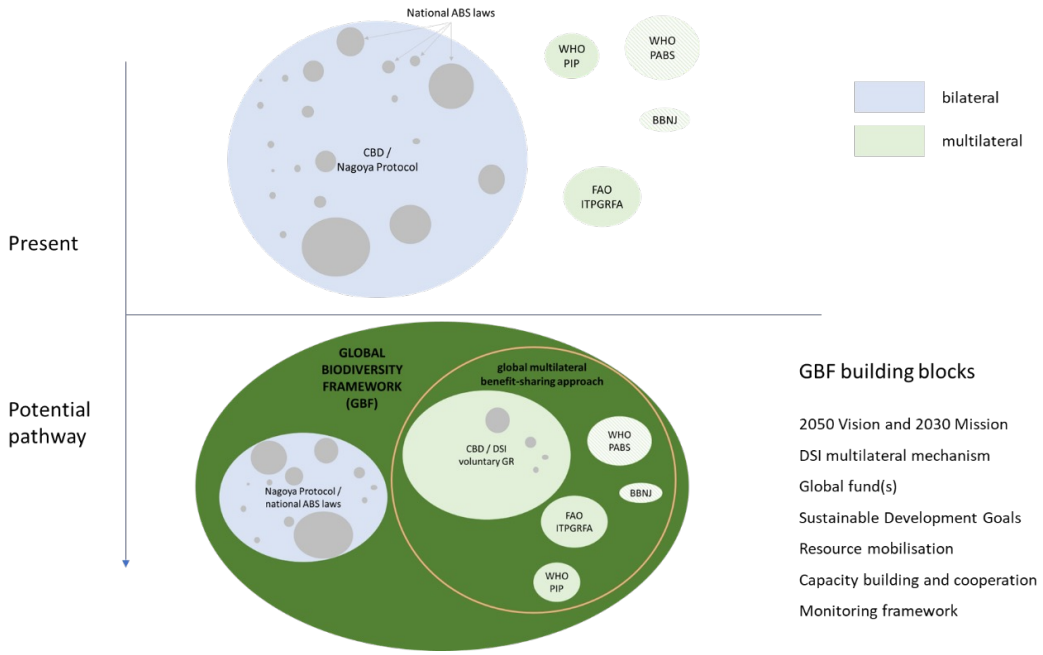

Figure 1: A visualization of a possible way forward for a harmonized system for benefit-sharing of DSI. Courtesy of Thomas Vanagt[1]

A harmonized system across all four relevant instruments needs to satisfy the following conditions:

- Mechanisms for benefit-sharing from DSI should be developed according to the existing scientific research infrastructure and practices, recognizing that all DSI data contained in thousands of biological databases and exchanged daily with millions of users can potentially be of relevance under more than one international instrument.
- Terms of use and trigger points for benefit-sharing for all 4 UN fora (CBD, WHO, FAO, BBNJ) should be harmonized from the user perspective so that they are able to access and use data without undue transaction costs, and ideally, with none. Simple user compliance, which covers broad obligations under all fora and provides overarching legal certainty, is needed.
- Access and use of DSI should be decoupled from benefit-sharing so that researchers can use all types of DSI for a variety of purposes, such as conservation, food security and health, regardless of which forum they are associated with.
- Open access should continue to guarantee that all users are able to access, compare, and analyze data across all databases and datasets. The system should also account for emerging tools and technologies, such as artificial intelligence, which might reduce the need to directly access individual sequences in deriving research outcomes and future benefits.
- DSI benefit-sharing mechanisms should be coherent across fora in terms of the kind of monetary and/or non-monetary benefits they expect from the use of DSI. Monitoring schemes could pool and maximize outcomes across all fora, while still supporting the specific goals of the various frameworks.
- Capacity-building to optimize DSI access and benefits should be broadly defined, conducted,

and accounted for with the aim to reduce the gap of capacities across different fields, such as conservation, food security and health, around the world.

If each UN forum develops divergent benefit-sharing mechanisms for DSI without considering how researchers use DSI, particularly if these uses fall simultaneously within the scope of two or more international instruments, negative impacts will result for both research and benefit-sharing.

The DSI Scientific Network calls on negotiators to take these critical issues for scientific research into account now, while different mechanisms are still under development. Such concepts and aspirations as OneUN and OneHealth will be unattainable if siloed DSI benefit-sharing systems are created. DSI is fundamentally a cross-cutting resource in science, and mechanisms regulating its use must be coherent and mutually supportive.

Case study #1: Using comparative DSI to determine if cold-active enzymes can be used in washing detergents at colder temperatures to reduce energy consumption

Cold-active enzymes have been described since the 1980s and have been associated with microorganisms living in temperature ranges of -2 to 20°C in a variety of environments such as deserts, mountains, wetlands, polar regions and deep-sea waters. They are ideal candidates for a variety of industrial applications, including as efficient enzymes used in laundry detergents to improve fabric protection and reduce energy consumption by washing at lower temperatures.

Arctic Sea Ice
Credit: Jacob Skowronek

The comparison of new sequences against DSI already publicly available in databases is a critical step for characterizing genetic material from novel organisms. The identification of the cold-adapted amylase enzyme, amy175, using DSI analysis and functional experiments, was a breakthrough for researchers seeking to develop more efficient and environmentally friendly detergents[2].

To identify and characterize the microbe that produces this enzyme, and the enzyme itself, researchers looked at hundreds of sequences from cold-active enzymes, microbes and a range of other sources. The research focused on function (i.e., whether enzymes could be efficient in cold water), not on geography or type of organisms, since these dimensions would be irrelevant to the qualities that researchers were seeking. In the end, fourteen sequences from the global databases (INSDC) were used

to confirm the microbe producing the enzyme, then twenty-seven amylase protein sequences from a specialized database were used to confirm enzyme family and function. The DSI used was from a broad range of organisms, including humans, frogs, insects, and bacteria.

This study shows how researchers are using data from a broad range of sources and across databases. DSI that could fall in the scope of BBNJ and CBD, as well as unknown DSI were used in various, dynamic proportions through the R&D process. The researchers working on this project would not know whether they should share benefits through a BBNJ or CBD framework. And some of the DSI could not be traced back to the origin, so compliance based on origin would not be possible.

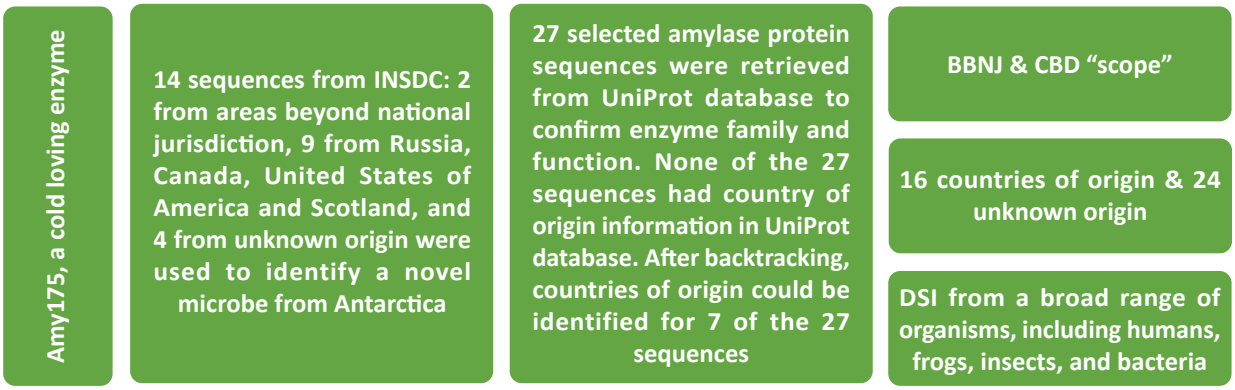

Figure 2: Research on cold active enzymes used DSI from a great variety of sources, which would be in the scope of at least two different policy instruments on ABS for DSI.

Case study #2: Using DSI to improve pathogen resistance in crops

Wheat is an essential staple food for one third of the world's population. 70% of the global production of wheat is used for human consumption, but pests and diseases are a growing threat. For example, stem rust, caused by *Puccinia graminis (Pgt)*, which had been well-managed through improved varieties since the mid-20th century, has seen a resurgence in recent years due to the emergence of new races of *Pgt*, particularly affecting the Middle East and Africa[3]. Another pathogen, *Zymoseptoria tritici*[4] , which causes Septoria tritici Blotch (STB) disease in wheat, has become a serious concern in Europe and North America as it started to develop resistance to the fungicides commonly used to manage STB. To prevent crop losses and support food security, research into new varieties of wheat with resistance to these diseases is a priority.

A recent study[5] in France looked at fungal and bacterial communities of three varieties of wheat (two susceptible to STB and one with a resistance gene, Stb16q) at three time points of the growing season. Using open access DSI databases (e.g., SILVA, UNITE), the authors were able to identify and compare fungal and bacterial communities in infected and healthy wheat leaves and infer whether their compositions could act as beneficial biocontrol against *Z. tritici*.

Ug99 stem rust on wheat spike
Credit: CIMMYT

Another study[6], focused on *Puccinia graminis (Pgt)*, sought to understand the evolutionary history of the Sr22 gene, found in a wide variety of plants and grasses. After looking over 80 plant species and performing DSI analysis, the researchers determined that plants with gene clusters might be more resistant than those with a single locus.

Zymoseptoria tritici	*Puccinia graminis*
Leaf Varieties: Annex 1, International Treaty on Plant Genetic Resources (IT-PGRFA)	**Crop samples:** Annex 1, International Treaty on Plant Genetic Resources (IT-PGRFA)
Pathogen samples: CBD framework	**Pathogen and other plant samples:** CBD framework
A total of 72,000 fungal and bacterial sequences from the SILVA and UNITE databases, without information on country of origin	DSI sequences from over 80 plants

Figure 3: Overview of the DSI used in two studies on wheat pathogen, demonstrating the relevance to multiple possible benefit-sharing frameworks and the great amount of data used.

In both studies, researchers used DSI from various genetic resources in a dynamic way throughout the R&D process. In the case of Z. tritici, all samples were taken in France, a country with a legal framework for benefit-sharing of genetic resources under the Nagoya Protocol, and a signatory to the Plant Treaty. As the geographical origin of the fungal and bacterial sequences was irrelevant to the research, this information was not available. Trying to infer this information for over 72,000 sequences would be a huge burden. For Pgt, similarly, DSI relevant to plants under ITPGRFA and to plants and pathogens under CBD were used, in addition to unknown DSI. Both these examples show how R&D involves accessing and using genetic resources and DSI that are regulated under different international instruments. And it highlights the negative potential impact on such research if the benefit-sharing mechanisms for related DSI adopted under those instruments are not coherent, undermining open access and interoperability of databases, creating additional burden on users of DSI.

Acknowledgement: This policy brief was supported by the German Federal Ministry of Education and Research (BMBF) WiLDSI VorweRts 16LW0062K

Literature cited

[1]*Hartman Scholz, A., Humphries, F., Vanagt, T., Jaspars, M. (2023) A new dawn for global benefit-sharing: capitalizing on the Global Biodiversity Framework for Marine Genetic Resources from Areas Beyond National Jurisdiction - https://www.iucn.org/sites/default/files/2023-02/bbnj_icg5bis_policy_brief_global_benefit_sharing_1.pdf*

[2]*Wang X, Kan G, Ren X, et al. Molecular Cloning and Characterization of a Novel α-Amylase from Antarctic Sea Ice Bacterium Pseudoalteromonas sp. M175 and Its Primary Application in Detergent. Biomed Res Int. 2018:3258383. Published 2018 Jun 27. https://doi.org/10.1155/2018/3258383*

[3]Singh, R.P., Hodson, D.P., Jin, Y., Lagudah, E.S., Ayliffe, M.A., Bhavani, S., et al., 2015. Emergence and spread of new races of wheat stem rust fungus: continued threat to food security and prospects of genetic control. Phytopathology, 105, 872–884. https://doi.org/10.1094/PHYTO-01-15-0030-FI

[4]Torriani S.F.F., Melichar, J.P.E., Mills, c., Pain, N., Sierotzki, H., Courbot, M., 2015. Zymoseptoria tritici: A major threat to wheat production, integrated approaches to control. Fungal Genet. Biol. 79, 8–12. https://doi.org/10.1016/j.fgb.2015.04.010

[5]Barroso-Bergadà, D., Massot, M., Vignolles, N., d'Arcier, J.F., Chancerel, E., Guichoux, et al. 2023 Metagenomic Next-Generation Sequencing (mNGS) Data Reveal the Phyllosphere Microbiome of Wheat Plants Infected by the Fungal Pathogen Zymoseptoria tritici. Phytobiomes Journal 7:2, 281-287. https://doi.org/10.1094/PBIOMES-02-22-0008-FI

[6]Hatta, M. A. Md., Ghosh, S., Athiyannan, N., Richardson, T., Steuernagel, B., et al. 2020. Extensive Genetic Variation at the Sr22 Wheat Stem Rust Resistance Gene Locus in the Grasses Revealed Through Evolutionary Genomics and Functional Analyses. Molecular Plant-Microbe Interactions 33:11, 1286-1298. https://doi.org/10.1094/MPMI-01-20-0018-R