



# What does open access to DSI mean for researchers?

Microscopic images, like this cross-section of plant cells, are used in combination with open DSI to connect the dots and generate new knowledge for biodiversity.

**Digital Sequence Information (DSI)<sup>0</sup>** is mostly available through public databases where it can be accessed openly by researchers worldwide. This “openness” is what allows the broader DSI ecosystem to exist and function. The value of DSI comes not from access to single sequences, but from researchers’ ability to sift through the millions of sequences available in the global dataset and compare their data against these.

**Open Access to DSI is crucial because it allows scientists to generate knowledge that helps to protect biodiversity, advance sustainable development goals and address global health crises.**

**Open access to DSI drives research and innovation** by lowering barriers to access to information and facilitating international research collaborations. It also promotes scientific integrity and enables reproducibility.

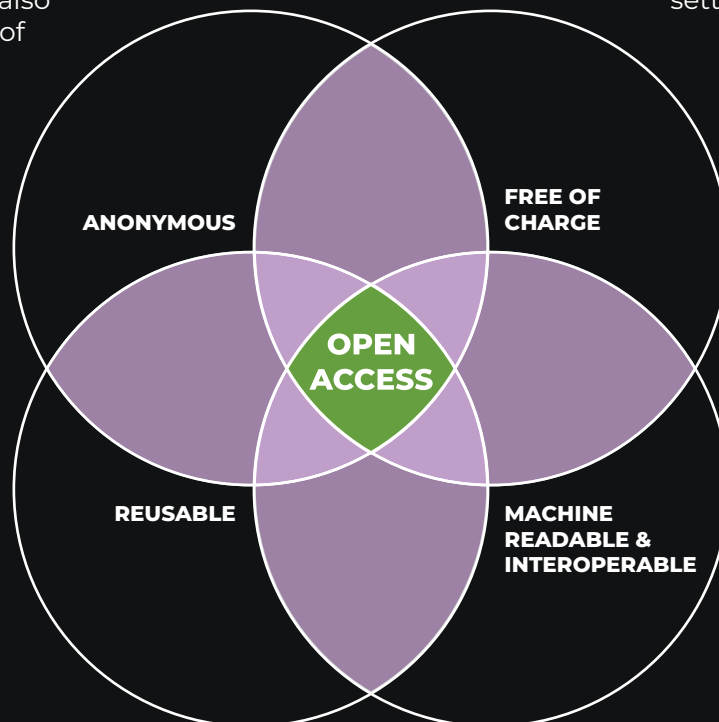
# What do we mean by “open access”? While there is no set definition, for most researchers using DSI, there are four key interdependent qualities of importance to an open system:

## ANONYMOUS

Most public DSI databases can be accessed without registering or creating an individual login. This allows researchers worldwide to access and use the dataset, independent of their institutional affiliation, and reduces costs by eliminating the need for a system to manage user accounts. It also simplifies automated use of the data through remote interfaces, by enabling a seamless exchange of information across thousands of different databases.

## FREE OF CHARGE

Access to and use of public DSI databases is free of charge, regardless of the purpose or outcome of the data use. This means that there are no financial barriers in place, allowing all researchers across the world to access and use the dataset, whatever their resource-setting. This is key to help create a level playing field for research.



## REUSABLE

Researchers are able to reuse data accessed through public DSI databases without restrictions. If data cannot be reused, its contribution to the generation of further knowledge is limited. Reusable data allows researchers to save costs, time and labor. It is also necessary for researchers to be able to re-evaluate and reproduce research, in support of scientific transparency and integrity.

## MACHINE READABLE & INTEROPERABLE

Data is provided in a form that can be easily processed, transferred, analyzed and reused through a computer and across different systems. It is important for the data to be machine-readable as searches are not done manually by researchers, but largely run using applications and software. Data also needs to be interoperable. This means databases can “speak” to each other and data in different databases can be read just the same. Interoperability enables cooperation between researchers and between databases.



As part of this open access system, the scientific community encourages transparency and reproducibility through the attribution and recognition of previous scientific work and clear attribution to the work of others.



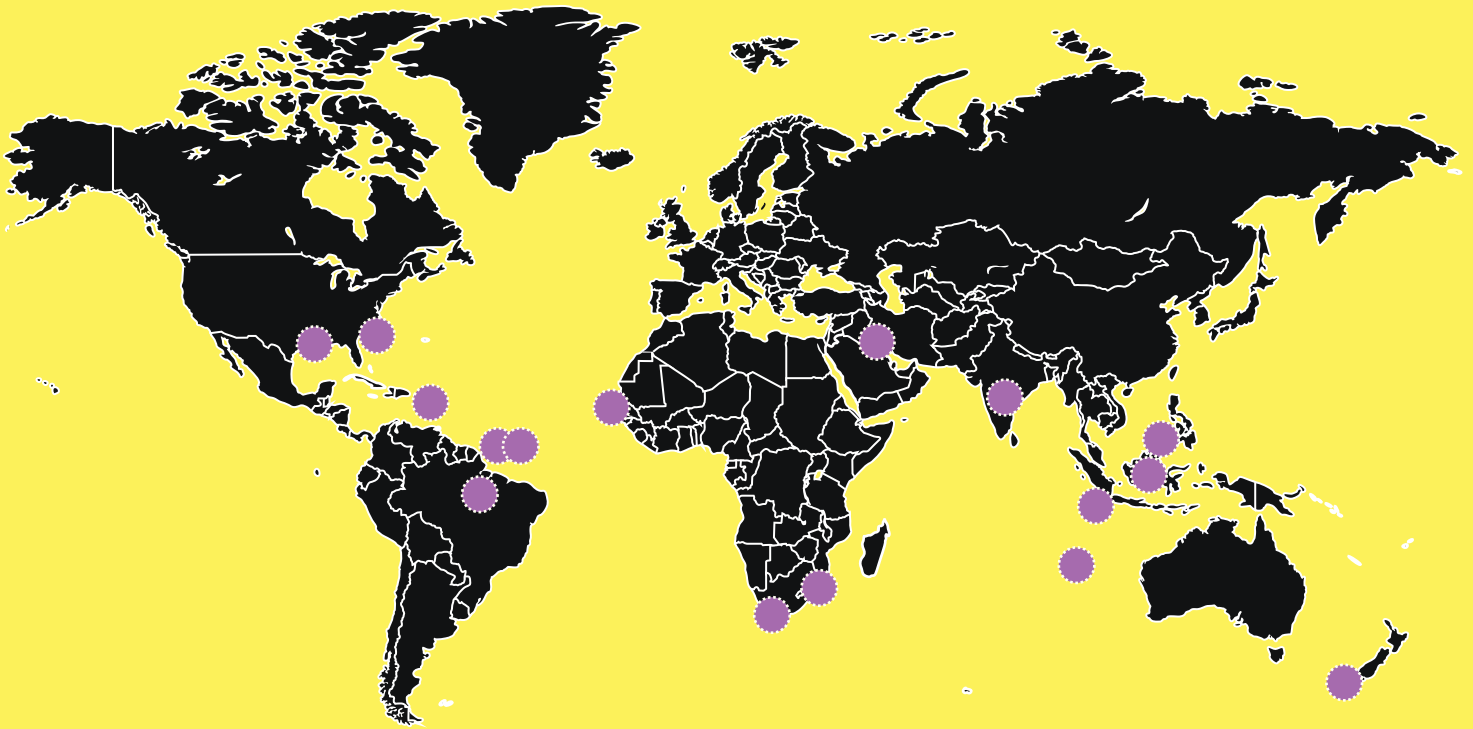
## Leveraging open access to DSI to protect endangered shark species<sup>3</sup>

Illegal wildlife trade poses a severe threat to biodiversity and animal welfare globally. In Brazil, endangered shark species are being fished for food consumption despite existing regulations which forbid their trade. Because the sharks are brought to land and sold without their heads and fins, it is difficult to identify illegal catches once they reach the market.

A group of researchers from Brazil and Australia were able to leverage information available in open-access datasets to address this challenge. They first extracted DNA from over 400 suspected shark meat samples collected across fish markets and local fisheries on the northern coast of Brazil. They then compared specific sequences from these samples with the entire nucleotide database of the NCBI (GenBank) and the Barcode of Life databases, which allow free of charge and unrestricted reuse of stored sequences.

By processing, analysing and reusing the sequences available in these vast public datasets, the authors were able to identify 17 shark species among the samples they had collected, of which 4 are listed as near threatened and 9 as threatened. These included species such as the great hammerhead shark, proving that endangered shark species were traded. To analyse the relatedness of the identified species, the authors used specific sequences from the NCBI nucleotide database, obtained from shark samples collected from around the world.

The development of these DSI-based identification methods can help policymakers quickly identify illegally-traded species and implement measures to protect them. This study illustrates the value of open access to DSI for biodiversity conservation and for achieving the targets of the Global Biodiversity Framework.



*DSI from the NCBI nucleotide database, obtained from shark samples collected from around the world, was used to analyze the relatedness of different shark species identified as part of this study.*

# Tackling food insecurity through open access to DSI: rice bacterial blight<sup>4</sup>

Rice is the world's most important food crop. It is the main source of calories for over half of the world's population. However, across many Asian and African countries, rice yields are threatened by a deadly bacterial disease: rice bacterial blight. The disease is caused by the bacterium *Xanthomonas oryzae* pv. *oryzae* and is among the most destructive for rice crops.

To tackle this issue, an international team of researchers reused previously published sequences from African and Asian strains of the bacterium, which are available under anonymous, free of charge and unrestricted access in the GenBank database. The authors also generated sequences themselves and deposited them in the database.

Thanks to the datasets being available in a machine-readable and interoperable format, researchers were able to easily process the sequences to compare the relatedness of the different strains. They were also able to analyze their genomes and extract information on disease-inducing genes. This process ultimately allowed researchers to generate rice lines that can resist the disease.

This study highlights that by ensuring open access to DSI, we can help fight devastating plant diseases and ensure food security.



DSI from different African and Asian strains of the bacterium *Xanthomonas oryzae* pv. *oryzae* were used as part of this study.

## NOTES

- 1 Digital Sequence Information, or "DSI", is a policy term that refers broadly to genomic sequence data and other related digital biological data. This includes the details of an organism's DNA and RNA, which determine its characteristics and unique traits.

## REFERENCES

- 2 Sara, R., Hufton, A. L., Sett, S. and Scholz, A. H. Open access: a technical assessment for the debate on benefit-sharing and digital sequence information. *Zenodo* (2022). <https://doi.org/10.5281/zenodo.5849643>  
Sara R., Hufton, A.L. and Scholz, A. H. Compatible or Incompatible? DSI, Open Access and Benefit-sharing. *SocArXiv* (2021). <https://doi.org/10.4324/9781003301998>  
Jaspars, M. and Brown, Abbe E. L. What Should We Mean by 'Open Access'? in C Lawson, F Humphries & M Rourke (eds), *Access and Benefit Sharing of Genetic Resources: Problems and Solutions*. Routledge (2022). <https://doi.org/10.4324/9781003301998>
- 3 Feitosa, L.M., Martins, A.P.B., Giarrizzo, T. et al. DNA-based identification reveals illegal trade of threatened shark species in a global elasmobranch conservation hotspot. *Scientific Reports* (2018). <https://doi.org/10.1038/s41598-018-21683-5>
- 4 Oliva, R., Ji, C., Atienza-Grande, G. et al. Broad-spectrum resistance to bacterial blight in rice using genome editing. *Nature Biotechnology* (2019). <https://doi.org/10.1038/s41587-019-0267-z>



This infographic was developed by the DSI Scientific Network with the generous support of the Norwegian Agency for Development Cooperation (NORAD). The DSI Scientific Network is a global network of researchers speaking out for sensible policies on DSI. For more information, visit our website: [www.dsiscientificnetwork.org](http://www.dsiscientificnetwork.org).

