3 open access scenarios for DSI in the framework of the CBD

A discussion paper for the European DSI scientific stakeholder workshop

Organized by the BMBF project "Science-based solutions for DSI" (WiLDSI)

Co-hosted by the Horizon 2020 European Virus Archive project (EVA-GLOBAL)

March 10-11, 2020, Brussels, Belgium

Background & Project description

In October 2020, at the 15th Conference of the Parties (COP15) to the Convention on Biological Diversity (CBD) a Post-2020 Global Biodiversity Framework and the issue of "digital sequence information" (DSI) from genetic resources and access and benefit sharing thereof will be negotiated. The contents of public databases for DSI are growing exponentially and some countries of origin fear that uncontrolled access to freely available DSI undermines the rightful sharing of benefits from the utilization of genetic resources. At the same time, DSI and its free accessibility are essential for life science, including biodiversity research, food security, human/plant/animal health and beyond.

The goal of the WiLDSI (Wissenschaftsbasierte Lösungsansätze für DSI) project is to provide scientific input prior to the negotiations and expected political compromises that could have long-reaching consequences for the global scientific research community. The WiLDSI project builds off of a prior workshop in Bonn in January 2020 and knowledge gained from research for the CBD Secretariat on DSI databases and traceability [1]. We assume the reader understands the DSI core infrastructure (INSDC) and downstream public and private databases. At the outset of the project we focused our research on three guiding principles:

1. Open access [2] is critical – not simply because it is "easier" for scientists – although the ease of data flow is a key element – but because it is part of the EU and Member States' long-term strategies as demonstrated by large financial and infrastructure commitments, the FAIR principles [3], and the European Open Science Cloud (EOSC)[4]. Scientific colleagues in the Global South widely support open access [5] and there is growing traction that the open access model should be preserved. However, undoubtedly some changes to the status quo will be called for. The question becomes what can be changed and at what cost to users and providers? Where are the boundaries of open access? We know as scientists that open access may "break" if pay walls are established and sometimes even if registration is required but where exactly are the boundaries of open access for our community?
2. Monetary benefit sharing will be required. Indeed the word "open", at first glance, seems to stand in direct contradiction to an income-generating system, which is a clear demand from the Global South. However, open does not equal free of any obligations; models can be deployed where DSI are visible to all, yet certain types of use of the DSI may be subject to conditions. The questions here are how to generate income without closing off access or causing high transaction costs. The research developed in the project started with an open system and learned from innovative financing models from other sectors with the goal of enabling fair monetary benefit sharing without compromising scientific integrity.

3. Non-monetary benefits need to be quantified, valued and communicated. The core DSI infrastructure, the International Nucleotide Sequence Database Collaboration (INSDC) is a permanent open access platform for DSI. It provides enormous societal good, connects thousands of scientific databases and platforms, and helps to level the playing field around the entire globe precisely because it is free and open. Yet it costs upwards of $50 million annually and is financed by the USA, mainly European nations [6], and Japan. The majority of the DSI in the databases comes from (is geographically sourced from) these countries and China and Canada and not from so-called "provider" countries in the global South [1], which inverts the expected provider/user dichotomy. There are 15 million users from every country in the world half of whom live in countries that do not pay for the infrastructure. In addition to the INSDC, large publication databases connect open access literature to the scientific results and datasets found in the INSDC and the surrounding infrastructure. Further DSI infrastructures such as Global Biodiversity Information Facility (GBIF), International Bar Code of Life (iBOL), Global Genome Biodiversity Network (GGBN), and many others build up this non-monetary benefit system that directly contributes to the first two goals of the CBD. The use of this infrastructure can be globally quantified, and results used in international discussions.

In this context, it is important to note the INSDC is a scientific collaboration between the European Molecular Biology Laboratory (an inter-governmental treaty organization), the National Center for Biotechnology Information (part of the US government) and the DNA Databank of Japan (part of the Japanese National Institute of Genetics), is under no *formal* obligation to make any changes in response to requests from the Convention on Biological Diversity (CBD). Furthermore, given that the majority of the data holdings and database access from INSDC databases do not come from CBD-relevant sequences, a reluctance to change INSDC operations is anticipated. At the same time, CBD-relevant DSI adds scientific value and the INSDC is motivated by a spirit of collaboration and, within reason, will work towards solutions that preserve data openness.

In September 2019 we began expert-led research on: potential monetary benefit sharing mechanisms, possible adaptations that could be considered by the INSDC, technical and policy implications of traceability, and models for open, non-traceable (de-coupled) systems. The confluence of these ideas has led to the development of the three scenarios below for how DSI could be handled in a new policy framework**. The goal of this workshop is to test the below scenarios, examine feasibility, propose new adaptations, and refine accordingly.** The two latter scenarios below focus solely on DSI, but could also be extended to the *physical* genetic resources, for example, by a "coalition of the willing". In particular, genetic resources held by public collections which could be distributed under standard terms and conditions seem well-suited to multilateral systems.

In November 2019 a first global ABS Dialogue was hosted by South Africa and Norway in Pretoria and led to the development of five DSI options [7, see Appendix]. To connect our results to international discussions, we have mapped our research onto the Dialogue options in the scenarios below. Of the 5 Dialogue options, we focus on options 2-4 and do not discuss 1 and 5 because, respectively, they do not ensure open access and monetary benefit sharing. While we firmly believe that 5 is important and should be emphasized (see goal 3 above), we believe that capacity building complements any scenario but is insufficient as a stand-alone option. The results of this workshop along with quantitative data on non-monetary benefit-sharing, data from interviews with international scientists, and INSDC policy assessments will be published in a report in April on the WiLDSI project website [10].

A) Nagoya plus (Dialogue Option 2)

DSI Handling: Although some countries already regulate DSI in their national legislation, one of the most simple and practical problems for a scientist seeking to comply with Nagoya-related obligations faces, is that it is not technically possible to know which sequences in the public databases have ABS obligations because legal documentation is not linked. A second problem is that even if it were technically feasible to create a linkage between ABS obligations and DSI in the databases, ABS permits/agreements (PIC/MAT) from most countries are still not published as internationally recognized certificates of compliance (IRCCs) despite the text of the Nagoya Protocol requiring Parties to use IRCCs and the ABS Clearing-House (ABS-CH, Article 14.2(c)).

IRCCs have two important features: 1) they give users legal certainty because they are published by a verified authority on a legitimate site and 2) they have a unique identifier and stable website address that would enable linkages to other data forms such as DSI. (Paper/PDF documents issued by Parties that are not IRCCs have limited legal certainty and cannot be linked to other digital data in a standardized or stable manner.) Thus, if Parties to the CBD/NP were to *exclusively* rely on IRCCs for the utilization of genetic resources as well as DSI generation, production, and usage, it would create a technological basis for improved "connectivity" and transparency between digital data and genetic resources. Parties could adopt standardized/simplified terms and conditions [8] (similar to open source software licenses) that would make downstream user compliance more efficient. Both of these areas (transparency on ABS obligations in the sequence databases and universal use of IRCCs) should be addressed before any traceable DSI solution could be considered.

Benefit sharing & Payment allocations: The "Nagoya plus" scenario[a] would create a mechanism (using perhaps the ABS-CH) to link IRCCs with standardized conditions to DSI in the public databases using the Accession Number from INSDC thus creating a new linkage between ABS and DSI. Users of GR that submit DSI to the public databases would be responsible for creating this ABS connection by adding the IRCC number during the DSI submission process to the INSDC. Subsequent users that access the database and use DSI would be required to check the conditions of the IRCC before use. Benefit sharing would take place on a bilateral basis similar to the Nagoya Protocol. While performing compliance checks, checkpoints could inquire about the DSI AN(s) used during utilization and the accompanying IRCC(s).

Changes required:

- Parties: Commit to issue only IRCCs if they claim sovereign rights over DSI and to standardize terms and conditions of DSI-based IRCCs. Alternate forms of PIC/MAT not published through the ABS-CH would be legally interpreted as waiving a Party's claim to DSI benefit sharing. ABS-CH would need to be adapted for linking IRCCs and DSI.
- Users: Obligation to add/link ANs to IRCCs using the ABS-CH. Downstream users would need to check ANs in the ABS-CH to determine if benefit sharing obligations exist.

[a] In the Dialogue report [7] and accompanying graphic, option 2 uses a country tag. Here, we have replaced country tag with an IRCC since a country tag alone would not explain to the user what the benefit sharing obligations are. If a country tag were to replace the IRCC in this scenario, then there would need to be a reliable, legitimate website for each country explaining how DSI is handled and what the standard terms and conditions for the country are. In either scenario, the user would still need to keep track of which DSI was used.

- Monitoring: The checkpoints responsible for compliance checks according to Article 6 of the Nagoya Protocol also check users of DSI by requesting Accession Numbers (ANs, unique identifiers issued by INSDC) and IRCC numbers during compliance checks.

Pros:

- Relatively simple changes that build on and complement the existing bilateral system that has been in development since 2010. Significant capacity development has already been invested here.
- Recognition and coupling to physical GR where DSI always originates from.
- User familiarity with the ABS-CH has grown and would continue to grow over time.

Cons:

- Significant time delay until monetary benefits begin to accrue and likely low monetary outcomes as judged by current frustrations with the implementation of the Nagoya Protocol [9].
- There is potential for heterogeneity of different use conditions and ABS obligations. Modern biological research is based on huge datasets where thousands or even millions of DSI are used which makes the consideration of all terms and conditions and thus implementation of the system very challenging.
- This scenario over-simplifies the complexity of the flow of DSI between scientists, databases, publications, and other experimental data types. There are >1,700 downstream databases that pull, parse, transform, annotate and modify DSI out of the INSDC. It would be challenging for downstream databases to adequately capture and display legal information. Thus, it is likely that some or many of the >1,700 downstream databases will chose not to import DSI data associated with IRCCs regardless of the conditions described therein; this is because of the complexity of integrating data sets, and serving these to onward data users, where the data sets are associated with different conditions of use.
- Biology is redundant. There are millions of repetitive DSI entries or parts of entries in the databases meaning that even for biologically novel sequences, it is likely that an alternative sequence from a non-IRCC-linked sequence entry could be found. And given that a least half of the ISNDC DSI dataset is entirely out of CBD scope [1] (human, model organisms, biodiversity from non-party or free access countries), jurisdiction shopping – selecting DSI based on an absence of ABS conditions – will occur.
- It would be difficult to calculate the actual value of a particular sequence in a final product. While it might be possible to define certain criteria for such calculation/valorization, some data processes, such as the “BLAST” search, which requires a search database of all available sequences, might so dilute sequence contributions, that they should be considered exempt.
- It is unclear how checkpoints could monitor compliance with benefit-sharing obligations at the point of commercialization.

B) Country tag (Dialogue option 4)

A system that acknowledged the country of origin but handled the benefit sharing obligations in a multilateral system would remove the need for tracing DSI through the database infrastructure and through downstream (re-)use. This scenario would use the geographical information and policies of

the INSDC infrastructure as tools to standardize benefit sharing obligations and provide a way to recognize and reward the contribution of a provider country to the INSDC infrastructure all while maintaining open access for scientific use.

DSI Handling: DSI in the public databases can already be labeled with a "country tag" (used currently as scientific metadata not legal compliance) filled out in during the DSI submission process. However, around 43% of INSDC entries do not have a country tag *although they should have had one* [1] (determined by a manual review of publications associated with non-country-tagged DSI).

In this scenario, the INSDC would stringently ensure that the country of origin is reported during the submission process, curate missing country tags, and, together with the CBD Secretariat or Parties, organize an awareness-raising campaign to increase scientists' compliance. In parallel to this change in country-tagging practice, the **use policy** of the INSDC could be adapted to require a standardized contribution to a monetary benefit-sharing fund upon commercial use of DSI.

Benefit sharing: Commercial use of any CBD-country-tagged DSI or the entire INSDC dataset (including for BLAST or ftp download of the entire dataset) would trigger benefit sharing in a subscription-like contribution manner. The user would "click" their acknowledgement of this policy during access of INSDC but would not be required to register. If the updated use policy indicated that use includes not only individual users but also the downloading/interconnected databases (the 1,700 databases that are built around the INSDC and download and network with the INSDC architecture), which then obligated the downstream databases to pass on monetary benefit sharing obligations to their users, the benefit sharing mechanism could have a large reach across the biological database landscape since the INSDC plays a core role -- 99.9% of studied databases are connected to the INSDC -- and thus promising opportunities for significant monetary benefit sharing. Again here non-commercial use would be exempted.

Commercial use determination could be based on a using entity's tax status or its commercial intent. Users would be legally obligated to an annual up-front monetary contribution based perhaps on a percentage of annual profits which would go into a general fund. Importantly, *access to DSI would remain open* but the use of the data would be under a new use policy.

Enforcement of this use policy could appear weak – how do you know that commercial users of DSI will comply with the terms listed in the INSDC use policy? This type of open model is already in use in the open source software world. Open source software use policies are routinely reviewed for legal compliance within the legal departments of for-profit companies because these companies require legal certainty before embarking on costly development projects. Because this system would have very low transaction costs with very high legal certainty, the likelihood of compliance could be quite high. Furthermore, the near universality of the INSDC in biological research makes it difficult for a commercial user to argue that DSI was not used over the course of product development especially since Rohden et al., [1] showed that all private in-house databases that were interviewed downloaded the INSDC dataset at regular intervals. As part of risk-based checks under the Nagoya Protocol, checkpoints could simply ask individual commercial users for proof of annual contribution to the multilateral fund which would give a "pass" for downstream commercial development with DSI.

Payment allocations: The funds could be distributed based on the contribution of DSI per country to the INSDC on an annual basis which could be assessed in partnership with the INSDC. This would

have the side effect of motivating countries to contribute and report their biodiversity thus supporting the first goal of the CBD. Although it would be important to ensure that DSI is novel and not just deep sequenced highly repetitive "junk" DSI.

Changes required:

- Parties: designate sovereign rights over DSI to the multilateral system. (If necessary, it is conceivable that this mechanism could be available for a "coalition of the willing" rather than an obligation for all Parties.)
- INSDC: stringently ensure use of country-tag in the public databases. Hire new staff to curate country information around DSI. Implement and provide information and updated use policy explaining the country tag and refine international data management procedures accordingly. Curation and screening of DSI submissions likely needed.
- Users: Commercial use (not only commercial users) would trigger an annual monetary contribution which would go into a general fund. Submitters of DSI would need to increase diligence in using the country tag as well as potentially other related metadata, for example, temporal information, such as access and utilization date.

Pros:

- An "early intervention" monetary benefit sharing mechanism enables short-term generation of income. This could improve and build trust between Parties.
- Harnesses the existing strength and openness of the data infrastructure.
- Could encourage Parties to contribute biodiversity to the INSDC if benefit sharing is tied to DSI contribution.
- Low transaction costs. Tracking and tracing not required.
- Better country/geographic information in the INSDC is good for science and CBD goals 1 & 2.

Cons:

- The role of the INSDC changes significantly including increased staff requirements for DSI curation (country tag, temporal information) and an altered role of legal compliance monitoring. Furthermore, it is unclear if the INSDC would be willing to alter their use policy.
- Jurisdiction shopping reduced but not eliminated.
- Depending on changes to use policy, some downstream databases might not "pull" country-tagged data into their databases as it raises complexity for their downstream users.
- Determination of appropriate monetary contribution and rates could prove challenging.

C) De-coupled (Dialogue option 3)

It would also be possible to generate monetary benefits with a "de-coupled" scenario where monetary benefits are triggered downstream at the point of commercialization and nearly all DSI is handled in a single manner.

DSI Handling: In order to simplify all the different legal parameters (temporal, geographical, material) this scenario would simply ignore the legal fact that roughly half of the content of INSDC is out-of-scope of the CBD. Here "de-coupling" means that legal scope is ignored to create simplicity. (Although the 12% of the INSDC that is of human origin would not trigger benefit sharing.) This would

relieve users, providers, and checkpoints from determining benefit sharing obligations by an IRCC, country-tagging, or any sub-division of the DSI dataset. This in turn means that tracing is not necessary and the user simply needs to ask whether non-human DSI (i.e. the biological diversity of some type of life on this planet) was used at some point during the commercial development process. Simply put, all non-human DSI in the INSDC or any biological database would be viewed as biodiversity, which, scientifically speaking, it actually is. As scientists, this idea is rational since life knows no borders, moves freely around the planet and, at the level of DNA, evolution continuously "recycles and reuses" sequences via evolution and continuously mixes DNA anew via asexual and sexual reproduction. For example, the in-bred model mouse is around 99% genetically identical to many wild mouse populations around the world and, of course, came from the wild at some point in time. So why should these two types of biodiversity really be handled differently?

Benefit sharing: Because the definition of bio-DSI is simple (all non-human), it is easier to delay monetary benefit-sharing if and until commercialization happens rather than to tie benefit-sharing to access. Thus "de-coupling" means that "ABS" becomes "A….BS" – benefits sharing is de-coupled and, in essence, delayed until benefits are ready to be shared. The user at the point of commercialization would contribute to a fund either based on a percentage of product sales (royalty) or a one-time payment fee based on the sector of the product. Here we foresee different commercialization fees for each sector. Commercialization would be defined as market entry and would vary based on the type of product. Market entry points are diverse but finite and regulatory oversight of many of these types of products could offer checkpoint opportunities.

However, because this scenario ignores the legal reality of DSI, we would suggest that this scenario be voluntary. The "scope" here is very broad – everything non-human; the contribution is late; the payment is voluntary. Importantly, although voluntary, the contribution provides legal certainty for all types of benefit-sharing obligations, conceivably even for multiple international fora. Because the Nagoya Protocol enables sovereign states to regulate DSI at their discretion there are already sequences in the INSDC that have benefit-sharing obligations. But it is extremely difficult to identify these (e.g., 43% don't even have the country tag correct). So in order for a commercial entity to limit its liability in this morass of uncertainty, a voluntary fund providing legal clarity and certainty once and for all, and potentially even for multiple fora, could be extremely attractive.

Payment allocations: All countries that claim sovereign rights over their DSI would be treated equally and entitled to apply for funds for biodiversity-related projects based on pre-defined criteria.

<u>Changes required:</u>

- Parties: designate sovereign rights to DSI to multilateral mechanism and provide legal certainty to users contributing to voluntary fund.
- Users: Early commitments would build trust in a voluntary system. Over the long-term, awareness would need to be raised across all bio-based sectors.

<u>Pros:</u>

- Direct link to commercialization outcome and benefits sharing
- No jurisdiction shopping possible
- Potential to extend to other international fora.

Cons:

- Late/slow income generation.
- Trigger for monetary contribution is somewhat unclear. “Market approval” can be difficult to define for some sectors. Although if the fund is voluntary, this becomes less problematic as this would be self-policing in order to gain legal certainty.

How to choose a monetary benefit sharing mechanism? In the two multilateral scenarios above, we have selected exemplary monetary mechanisms to provide a basis for discussion. However, realistically, there are multiple funding mechanisms and triggers that could be considered. In order to enable effective and sustainable benefit-sharing post-2020, we propose a three step process for assessing possible funding structures. As a first step, a needs-based assessment mechanism should be developed to identify what particular funding is required in which countries and at what level, be it at local research capacity, data infrastructure and connectivity, biodiversity monitoring, etc. Or perhaps the needs extend across the entire value-chain of the emerging bio-economy. A second step would be to design a tailored funding approach that would aim to deliver the funding needed to the relevant parties. This tailored funding approach can draw from a spectrum of solutions including a subscription model, public paying domain (PPD), micro-levy, impact bonds, open access publishing fees, and certification schemes. All of these can be complemented by a governance structure such as a public-private partnership. As a third step the proposed approach needs to be fine-tuned through discussions with a wide range of public and private partners to bring in not only private funds but also to optimize output, timing and risk-sharing tools.

One example of a needs-based assessment and targeting that would be compatible with both the **Country** and **De-coupled** scenarios would be to use the monetary fund to broaden and expand the INSDC infrastructure to additional international sites, thereby increasing capacity and overcoming internet bandwidth and connectivitiy challenges experienced in South America, Africa, and Asia. One could imagine adding INSDC mirror sites, for example, in China, Brazil, South Africa, and India where scientific training is strong and DSI production and usage is growing rapidly but where the infrastructure still lags behind demand. A historical model for this is the development of the CGIAR system in the 1970s during the Green Revolution which initially was based on contributions mostly by countries and donors from the Global North but eventually led to self-sustaining infrastructures and strong capacity building supported by a large number of countries including some from the Global South.

Final thoughts: As international discussion on DSI proceed, no individual scenario will be perfect. Before establishing a new regime based on hypotheticals, it could be interesting to take a scientific approach and experimentally test and compare favored scenarios by initiating pilot projects or perhaps modeling different scenarios. Another issue is “expectation management” – what are realistic monetary outcomes and do they match expectations? These kinds of assessments can build long-term trust if expectations are realistically managed early on.

Finally, both the **Country** and **De-coupled** scenarios could potentially be relevant for other international fora such as UNCLOS ABNJ, FAO ITPGRFA, and WHO pathogen discussions. We hope that the WiLDSI project can inform not only CBD discussions but the larger conversation around ABS.

References

1. Rohden, F., Huang, S., Dröge, G., and Scholz, A., (2020) Combined Study on Digital Sequence Information (DSI) in Public and Private Databases and Traceability. Secretariat of the Convention on Biological Diversity. Montreal, Canada. https://www.cbd.int/meetings/DSI-AHTEG-2020-01.
2. https://en.unesco.org/open-access/what-open-access
3. FAIR stands for Findable, Accessible, Interoperable, Reusable. https://www.nature.com/articles/sdata201618
4. https://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud
5. Prathapan, D., Pethiyagoda, R., Bawa, K., Raven, P., Rajan, P., and 172 co-signatories from 35 countries (2018).When the cure kills—CBD limits biodiversity research. Science 360 (6396), P1405-1406.
6. EMBL member states: Austria, Belgium, Croatia, Czech Republic, Denmark, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Israel, Italy, Lithuania, Luxembourg, Malta, Montenegro, The Netherlands, Norway, Portugal, Slovakia, Spain, Sweden, Switzerland, United Kingdom. EMBL associate member states: Argentina, Australia. https://www.ebi.ac.uk/about/funding
7. http://www.absinitiative.info/fileadmin//media/Events/2019/68_November_2019__Pretoria South_Africa/Report-First-Global-DSI-Dialogue-SouthAfrica-201911.pdf
8. Oldham, P. (2009). An Access and Benefit-Sharing Commons? The Role of Commons/Open Source Licences in the International Regime on Access to Genetic Resources and Benefit-Sharing. Initiative for the Prevention of Biopiracy, Research Documents, Year IV, No. 11. Available at SSRN: https://ssrn.com/abstract=1438027
9. Pauchard, N. (2017). Access and Benefit Sharing under the Convention on Biological Diversity and Its Protocol: What Can Some Numbers Tell Us about the Effectiveness of the Regulatory Regime? Resources, 6, 11; doi: 10.3390/resources6010011
10. https://www.dsmz.de/collection/nagoya-protocol/digital-sequence-information

Appendix: Options 2, 4 and 3 from the DSI Dialogue in Pretoria (Nov. 2019) [7].