Digital Sequence information, Open Access, and Sustainable Benefit Sharing: Scientific Input to International Policy Decisions

March 10-11, 2020, Brussels, Belgium

Venue: NH Collection Brussels Centre

The workshop was jointly organized by the BMBF-funded WiLDSI project, the Leibniz Institutes DSMZ and IPK Gatersleben (Germany) and co-hosted by the Horizon 2020 Project, European Virus Archive (EVA) and the EVA institutes DSMZ, National institute for Public Health and Environment (RVIM, Netherlands) and the Pasteur Institute (France).

WORKSHOP REPORT

Table of contents

Background and Brief Overview

The Parties to the Convention on Biological Diversity (CBD) will meet in Kunming, China to negotiate and define international biodiversity targets for the next 10 years.[1] In parallel to these discussions, the issue of benefit sharing from "Digital Sequence Information on Genetic Resources" (DSI) will also be negotiated. At stake in these negotiations for biologists is open access to sequence data via the large public databases and the open science and research system built around it. To avoid worst-case scenarios, science-based solutions for benefit sharing that do not endanger open access are needed.

The German Federal Ministry of Education and Research (BMBF) has funded an interdisciplinary research project (WiLDSI) led by the Leibniz Institutes DSMZ and IPK Gatersleben to research DSI policy options and engage the scientific stakeholder community. The project, "*Wissenschaftsbasierte Lösungsansätze für Digitale Sequenzinformation (WiLDSI)*"[2], aims to develop a scientifically sound concept for DSI which could potentially provide the German government and EU negotiators with a science-based tool kit for the upcoming negotiations. Over the past six months, the project has brought together database, legal, scientific, and financing experts to research possible policy options for DSI.

To engage the European scientific community in the political discussions surrounding DSI, a workshop on "Digital Sequence Information, Open Access, and Sustainable Benefit Sharing: Scientific Input to International Policy Decisions" was held on 10-11 March 2020 in Brussels, Belgium. The workshop was jointly organized by the WiLDSI project and above institutes and co-hosted by the Horizon 2020 Project, European Virus Archive (EVA)-GLOBAL and the EVA participating institutes DSMZ, National Institute for Public Health and Environment (RVIM, Netherlands) and the Pasteur Institute (France). The main objective of the workshop was to provide a platform where scientists could voice their opinion, discuss their views, and test the feasibility of the three preliminary DSI open access scenarios developed by the WiLDSI project.[3]

The DSI workshop was attended in-person by 47 participants and several dozen online (due to the emerging coronavirus pandemic) from different European research institutions and infrastructures, private sector research, and policy makers from across Europe.

Day 1

The workshop was opened by Dr. Jens Freitag, (Leibniz Institute IPK, Germany) and Dr. Mery Piña (Pasteur Institute, France). The workshop was conducted under Chatham House Rules and any discussion during the course of the workshop would not be attributed to individuals. Participants were encouraged to openly express their opinions and put forth their points of view so that the WiLDSI project can provide a voice to the European research community with respect to the political debate surrounding the DSI issue.

[1] This meeting has now been postponed from October 2020 to May 2021 due to the COVID-19 pandemic.

[2] For more information on the WiLDSI project see: https://www.dsmz.de/collection/nagoya-protocol/digital-sequence-information

[3] The background paper explaining the three scenarios can be found here: https://www.dsmz.de/collection/nagoya-protocol/digital-sequence-information/dsi-workshop-march-2020

Presentations

Overview of the CBD decision-making process surrounding DSI (Alicja Kozlowska)

Alicja Kozlowska (Policy Officer ABS, European Commission) gave an overview on the CBD processes surrounding DSI. She explained that the CBD in 1993 put in place new principles whereby genetic resources were no longer considered 'heritage of mankind' and that States have sovereign rights over their genetic resources. The CBD also introduced the concept of access and benefit sharing (ABS). The discussions on DSI are taking place in this context. She went on to explain how the Nagoya Protocol (NP) was agreed in 2010 in order to 'balance' the Aichi targets and to further advance the implementation of the 3rd objective of the CBD, ABS. However, she posed the question whether this 'Grand Bargain', whereby the industrialized countries were supposed to get transparent access legislation and the developing countries fair and equitable sharing of benefits, really work in practise?

Many amongst the developing countries say they are frustrated with the narrow scope of the Protocol and dissatisfied with the perceived insufficient (monetary) benefits generated by the Protocol. Many in the developed world express frustration with intransparent processes and delays in research. In recent years, DSI began to be perceived as a loophole/leak in the system – a way to circumvent the NP obligations and one perceived reason why few monetary benefits were being realised.

She described how the DSI issue was addressed at the last two Conferences of the Parties (COP13 and COP14) and explained the inter-sessional processes that were established and the upcoming tasks of the Ad Hoc Technical Expert Group (AHTEG) in March 2020 and the third meeting of the Open-ended Working Group on the Post-2020 Global Biodiversity Framework (OWEG 3) in July 2020.[4] Importantly, as next steps, the OEWG3 will proceed to discuss the results of AHTEG and would recommend a decision on DSI for COP15. She stated that DSI will ultimately be part of the post-2020 decisions and extensive knowledge is currently being gathered during this inter sessional period (2018-2020). Finally, she stressed that there are high expectations from the upcoming AHTEG and OEWG3 meetings but the highest expectation is from COP15. She concluded by noting that DSI is also currently being discussed by other fora such as the ITPGRFA, FAO CGRFA, WHO, BBNJ-UNCLOS and WIPO.

Open Access and the International Nucleotide Sequence Database Collaboration (Dr. Guy Cochrane)

Dr. Guy Cochrane (Head of the European Nucleotide Archive, EMBL-EBI) provided valuable information on open access public sequence databases, in particular, the International Nucleotide Sequence Database Collaboration (INSDC). He began by stating that "Science proceeds through layers" whereby data published by one group of scientists is picked up by another group of scientists who build up on it and take it a little further. This is primarily how science advances layer by layer. He presented an interesting analogy by comparing the data generated by Omics technologies and bioinformatics to an organic machine that runs autonomously. The input to the "machine" are DNA sequences that is processed in order to produce/generate knowledge (output). Scientists, software, secondary databases, data curation activities, laboratory research, and other activities process these sequence data. To continue to generate and gain knowledge, this complex 'life science machine'

[4] The third meeting of the Open-ended Working Group on the Post-2020 Global Biodiversity Framework (OWEG) has been postponed due to the COVID-19 pandemic.

needs to be protected and maintained. He then went on to emphasize the importance of open data with examples from endangered species, antimicrobial resistance, crop plant breeding and phylogenetic relationships between microbial species. He spoke about the history, organisation and structure of INSDC and the three main large sequence databases that belong to it: The European Molecular Biology Laboratory - European Bioinformatics Institute (EMBL-EBI), the DNA Databank of Japan (DDBJ) and GenBank, which is hosted by the U.S. National Center for Biotechnology Information (NCBI). He briefly described the data submission process and the significant scale (1 new dataset every 6 minute; $2x10^9$ sequences and $1x10^{16}$ base pairs of read data across $2x10^6$ taxa; 2,000 submitters, 10x thousands monthly consumers and 10x millions of monthly hits) at which the database operates as well as the enormous scale of worldwide usage of DNA sequences from EMBL-EBI. Additionally, he explained that the *Nucleic Acids Research* journal has been tracking bioinformatics databases for over a decade and their registry of public databases the most biological databases link to the INSDC. He also gave examples of this linkage with secondary databases such as UniProt, MirBase and SILVA.

He concluded that high productivity drives scientific progress (referring to the life science machine analogy) and enables ABS. The "life science machine" itself is an open, accessible interface for science making and a crucial part of the ABS system because of its worldwide use and accessibility. This productivity needs to be understood and maintained and the risks that could damage the productivity of the machine need to be described as well as the core features that need to be preserved and identify opportunities to improve it. Finally, he presented certain 'What if' scenarios that would negatively impact productivity and emphasized that that it is extremely critical that data remains visible and that the terms of use must be uniform across sequence types. Interestingly, he provided a scenario whereby the location information of the sequenced material could be scientifically enriching and potentially an opportunity to increase the productivity of the machine. He ended his presentation with real-time graphic depicting worldwide live hits on the EMBL-EBI web portal.[5]

Brief inputs from different scientific perspectives

Three scientists were invited to give short impulse talks (5 min.). The participants presented information regarding the usage of sequence data in their daily scientific life, the scientific questions they address as well the lessons learned from the Nagoya Protocol in this context.

1. Prof. Dr. Ibon Cancio, University of the Basque Country, EMBRC-Spain
2. Dr. Markus Wyss, DSM Nutritional Products Ltd.
3. Prof. Dr. Elżbieta Martyniuk, Warsaw University of Life Sciences.

Prof Dr. Ibon Cancio provided a brief description of his current research activities and emphasized the importance of open access of DSI in his research. His research involves: 1. the study of sex in fish present in the Basque estuaries with respect to the influence and effect of chemicals present in these estuaries, 2. the study of potent antioxidant molecule, Ovothiol A, in mussel ovaries. These research activities rely heavily on comparative analysis to homologous sequences of other marine invertebrates available in open access sequence databanks (such as Genbank). He briefly spoke about the annual ocean sampling day at the Plentzia marine station to study the microbiome made

[5] https://www.ebi.ac.uk/web/livemap/live-data-map.html

possible via sequence comparison to reference sequences available in the open access Ocean Microbial Reference Gene Catalog (TARA Oceans). He concluded his talk by emphasizing that the EMBRC marine stations are providers of research services; genetic resources (via Biobanks and culture collections) that is working towards following the Nagoya Best Practise Guidelines and data resources that is working towards open access, provenance models and FAIR data.

Dr. Markus Wyss spoke about the associated challenges if access restrictions were to be applied to DSI. DSM Nutritional Products Ltd is a science-based company in nutrition, health and sustainable living. It is involved in the field of animal nutrition and health as well as human nutrition and health, food specialities and personal care. He explained that products in the food and feed industry are typically composed of multiple components and that access restrictions could apply to any one of the individual components. Each component could potentially be derived from multiple GRs, and/or involving multiple "pieces" of DSI. How would one determine the value of each individual contribution? For example: Human milk oligosaccharides (HMOs) contain four different genes from four different microorganisms. With respect to DSI, DSM employs several production hosts (bacteria, yeast and fungi) for biotechnological purposes with whole genome sequencing of each production organism being common practise. These data are available free for further research and development. He also provided an interesting example of the generation of improved variants of commercial enzymes based entirely on DSI. For example, a thermostable phytase was developed based on sequence alignment of 153 freely available phytase sequences. He concluded by addressing the main CBD/NP challenges and their implications for typical food/feed products. These challenges include: 1. Issue of national sovereignty resulting in access restrictions and/or negotiations with multiple provider countries, 2. Stacking obligations and how to determine the value of each contributing component/sequence and assign "fair" share of benefits to each provider country, 3. Eternal obligations whereby over time the products will be further improved and the above issues would get worse over time and 4. No homology threshold with benefit sharing obligations even if used only transiently. These challenges represent a serious threat and it is very important to evaluate all options in order to generate a satisfactory and appropriate outcome.

Prof. Dr. Elżbieta Martyniuk provided a general overview of research conducted at the department of Animal Genetics and Conservation at the Warsaw University of Life Sciences. Their research involves working with: wildlife species (e.g. European bison, falcons etc.), dogs, sheep, chicken, laboratory animals, exotic species in zoos as well as animal origin products. Some of the scientific questions addressed include: searching for new alleles / genetic forms conditioning health-related traits (wild species), Phylogenesis, creating tools to identify a new breed (dogs), identification of genetic background of diseases (including idiopathic diseases, mutations in known genes, and development of molecular tests), genetic barcoding in fish species identification, identifying pathogens in animal origin products and detection of SNP (Single Nucleotide Polymorphisms) and INDEL (short Insertions and Deletions) variants for resistance / tolerance to Highly Pathogenic Avian Influenza (chicken). Their daily research activities also rely heavily on the use of sequence data with databases such as NCBI, Ensembl and Barcode of Life Data System (BOLD) being commonly employed. She spoke about lessons learned in the context of the NP describing certain issues related to the European bison project with Ukraine, Belarus and Russia. She concluded by emphasising the importance of open access DSI databases and if access to these DSI databases were to be limited it would lead to duplication of research, molecular tests as well as difficulty in obtaining reference sequences.

Update on the current COVID-19 outbreak with respect to DSI and the Nagoya Protocol (Dr. Mery Piña and George Haringhuizen)

Dr. Mery Piña (Pasteur Institute, France) gave an overview of the implementation of the Nagoya Protocol at the Pasteur Institute and the institute's response to the recent Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-COV-2) outbreak as a relevant case study. She explained that one of the main scientific areas of the institute is the field of emerging infectious diseases and, as such, the institute hosts 14 National Reference Centres (NRC), 9 WHO collaborating centres and has the 5th most diverse bacterial culture collection in the world. The institute's legal department assists in the implementation of the Nagoya Protocol providing appropriate legal counsel to the researchers and necessary follow-ups with the National Focal Points. The Pasteur Institute has a strong collaboration with the French ministries and policymakers in order to analyse practical implications of novel policies and determine its feasibility. Moreover, the Pasteur Institute has an international network comprising 32 member institutes that are spread over 25 countries. The institutes within this international network share their knowledge, research programs and provide on-going international health monitoring.

She then described the recent SARS-COV-2 outbreak and detailed the timeline and the work developed by the NRC for Respiratory Viruses during the SARS-COV-2 outbreak. The first confirmed case was detected in France on Jan 24th. A diagnostic test was developed using the Coronavirus sequences that were added in the Global Initiative on Sharing all Influenza Data (GISAID) database by the Chinese government highlighting the significance of open access information without which the scientific advances made during this current outbreak would have been virtually impossible. Interestingly, to date, the request to gain access to the Chinese virus isolate has been met with no response. The Pasteur Institute was the first in Europe to sequence the virus and the two full sequences have been submitted to the GISAID platform. On Jan 30th, a diagnostic test with 100% specificity for SARS-COV-2 was launched and on Jan 31st, SARS-COV-2 virus was made available by the Pasteur Institute for research purposes via the H2020 EVA-Global project. The institute is also sharing capacity building tools among the Institut Pasteur International Network (IPIN) members during this outbreak. Furthermore, a task force for COVID-19 was created on Jan 28th mobilizing Pasteur's scientific experts for the rapid development of diagnostic, preventive and treatment tools to tackle the novel coronavirus.

George Haringhuizen MA LLM (Coordinating Senior Counsel/Public Health Lawyer, RIVM, Netherlands) highlighted the importance of open access/open science by further elaborating and sharing his experience on the recent COVID-19 outbreak. RIVM is currently involved in the complex issue of negotiating the sharing of viral materials/sequences for producing antiviral compounds and eventually vaccines. This further highlights the need to share information openly and rapidly. Interestingly, the Coronavirus genomic sequence data seem to be better shared as compared to the actual physical material. He explained that if DSI was to be considered as covered by the scope of the NP, lessons would need to be learned from the current situation with respect to the sharing of the Coronavirus physical materials. He stated that the global sharing of physical samples of the virus were dramatically late or haven't happened yet. So countries had to wait for patients in their own country to become sick in order to obtain SARS-COV-2 samples. On the other hand, when scientists did share information/material, the receiving parties have no time and capacity for Prior Informed Consent (PIC) and Mutually Agreed Terms (MAT) negotiation due to the immense pressure for analysis, diagnostics tests, antiviral medication and vaccines. He noted that the recent COVID-19

outbreak requires international sharing of the viral sequence data and constitutes a key issue to control this outbreak and that sharing of COVID-19 sequence data has been comparatively better than in the previous Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS) outbreaks. He emphasized the need for speed on the sharing of genomic sequence data for two reasons: 1) rapid development of diagnostics tests/kits and 2) comparison of sequences (to develop phylogenetic trees) to study the transmission routes (epidemiology) of the virus and create response policies. If DSI was to be subjected to NP rules, this rapid sharing of sequence data that is extremely crucial during an outbreak would be impossible. He mentioned that scientists studying the Coronavirus are predominantly using the GISAID platform. Of note, there have been recent indications that the current rule of third party sharing of sequences from the GISAID database is being lifted in light of the recent events.[6] He concluded that there has been a massive non-monetary benefit sharing taking place among research institutes across the world through their scientific research, free performance of diagnostics, large-scale production of diagnostic kits with production partners against low/moderate prices, search for antiviral components and eventually vaccine production. At the moment, this large scale non-monetary BS is not being calculated/quantified or very visible. However, one must keep in mind if this were to be calculated or brought under MAT/MTA's, the people contributing to this BS for free may demand monetary compensation. This should be an important consideration if DSI was to be considered as falling under the NP regime. Lastly, he emphasized that it is extremely critical that the physical viral samples from the current COVID-19 outbreak be collected systematically and made freely available to scientists around the world to protect against the next outbreak.

CBD negotiations surrounding DSI: what is needed from different sectors? (Dr. Hugo-Maria Schally)

Dr. Hugo Schally (Head of Unit, Multilateral Environmental Cooperation EU DG Environment, European Commission) gave the participants a political perspective on the CBD negotiations surrounding DSI. The issue of DSI is complex and one must take into consideration the original framework of the CBD which recognizes the sovereign rights of States over their genetic resources. Genetic resources can have specific regulations applied to them at their point of origin and it is not difficult to imagine a claim that information that has been sequenced in digital form is therefore subsequently subjected to the rules of the country providing access. However, this does not necessarily imply that all information in the databases would be subjected to the NP. Taking some guidance from the previous participant impulse talks, the important points to consider are: predictability, legal certainty, clarity and uniformity of rules that are applied. There was a very clear understanding at the recent OEWG 2 (Second Meeting of the Open-ended Working Group on the Post-2020 Global Biodiversity Framework) in Rome (February 2020) that the discussion on ABS would be related to the broader concept of resource mobilization and the generation of financial resources that would eventually flow back to countries that are centres of biodiversity worldwide. This was the discussion surrounding Goal E of the Global Biodiversity Framework. Coming to the history of the NP, the Protocol was adapted in 2010, in the context of the Aichi Biodiversity targets, with the aim to achieve one of the three objectives of the Convention: the fair and equitable sharing of benefits

[6] Note: As the pandemic continued to grow after the March workshop, the GISAID database was soon unable to manage the volume of sequence data. The INSDC stepped in and created a coronavirus sequence data platform which has since become very widely used. https://www.covid19dataportal.org/

arising from the utilization of genetic resources. It was an agreement to provide a legal framework for genetic resources.

Unfortunately, prior to COP10, there was a lack of appropriate dialogue and in-depth discussions with the scientific research community and the industrial sector in order to understand better the contribution that the research community would have been willing to make and what solution would have best fit the needs of science and research in the context of an overall framework. After Nagoya, many questions were left open (e.g. genetic resources of uncertain origin) that eventually added to the frustrations of the developing countries. These countries began to assert that DSI/sequence data was the major reason for the absence of the flow of benefits. The biggest challenge therefore at the moment is to find a solution that: 1. Preserves open science, 2. Respects the legitimate concerns of the developing countries, 3. Establishes a system whereby benefits flow from users to providers. An important aspect is to include private sector stakeholders, as this sector is potentially the main source of the monetary benefits. Unless we come up with a workable solution, it is possible that we may end with another agreement which might bring DSI under similar compliance measures of the NP (PIC/MAT and traceability). The scenarios/options that have been developed and that will be discussed during the course of the workshop are creative enough if the scientific community is ready to adapt. In order to break out of the mold of the bilateral system that the NP offers, there needs to be readiness to engage in an alternative solution which may lie in the reflection on one of the options presented at this workshop. Finally, the scientific community needs to actively contribute to the political process in the lead up to Kunming, China next year.

Initiating the Dialogue: 'Fish Bowl' discussion

Participants put forth their questions concerning the political debate/negotiations surrounding DSI and ABS.

The main points addressed during this discussion:

- There are loose ends and open questions that remain from the Nagoya Protocol. Participants questioned whether closure and a new start are possible through a political solution.
- Due to time limitations and the complexity of the topic, many expect that an agreement at the COP15 (Kunming, China) on a system will not be possible, but rather that a framework could be agreed upon.
- A framework needs to link (some) research activities to commercial results thereby reflecting the contribution of DSI to the development of a product. However, this does not imply the need for a direct and fully-fledged tracking/tracing system but rather a direct link between a product and research activities that contributed to it.
- The Global Biodiversity Framework 2020 will try to repair what was not achieved by the Aichi Biodiversity targets.
- Policy makers would like to have an understanding of the 'business model' that the scientific community would be able to work with (with respect to the paradigm of open science and also in the context of the proposed scenarios). What are acceptable levels of change? What is the level of constraint that is acceptable?

- There is a need to better communicate the overwhelming complexity of tracking and traceability and the sheer volume of information in the databases.
- Resource mobilization and innovative finance mechanisms can play a more integrated role in ABS discussions.
- The economics of research and scientific data are not well understood. In other fields, data can be a monetizable commodity but research data does not fit this model.

The WiLDSI project & overview of the 3 open-access DSI scenarios (Dr. Amber Hartman Scholz)

Dr. Amber.H.Scholz (Deputy to the Director, Leibniz Institute DSMZ) presented a brief history of the WiLDSI project and described the three open access DSI scenarios that would be the main focus of discussion during day 2 of the workshop. She highlighted that the Leibniz Institute DSMZ is the first registered collection under EU Regulation 511/2014 and actively takes over a portion of the user's due diligence obligation. She explained survey results from biodiversity research showing that research projects conducted in NP countries with access regulations have average delays of 13 months and showed how the response rates of National Focal Points is often slow and incomplete and noted that this has caused significant frustration amongst biodiversity researchers. Furthermore, with respect to the current political dilemma surrounding DSI, Dr. Scholz explained how the WiLDSI project came into being and the four main reasons for engagement in the DSI policy process: 1) importance and exponential growth of sequence data, 2) acknowledgement that DSI can reduce but yet not replace access to GR, 3) biodiversity research and conservation need to use DSI and 4) lessons learned from previous CBD discussions. She went on to provide a crash course on the DSI core infrastructure: the INSDC and downstream public and private databases. In brief, this core infrastructure can be perceived as the nucleus from where the rest of the biological database landscape builds upon whereby 99.9% of the NSD databases depend on sequence data from the INSDC. Half of the data in the INSDC is out of geographical scope with 52% of the NSD being geographically-sourced from 4 countries: China, United States, Canada and Japan – i.e. *not* predominantly from low- and middle-income countries. Approx. one third of the INSDC is out of material scope (human sequences, sequences from model organisms etc.). She further explained some key lessons learnt from the CBD study on DSI databases and traceability.[7] Based on the understanding of the DSI infrastructure, the WiLDSI project came up with 7 project goals of what a new DSI ABS system should be:

1. Compatible with free, open access INSDC
2. Integrated with the existing database infrastructure used by science and NOT a stand-alone system (e.g., blockchain)
3. Administratively nimble or invisible for scientists
4. Ideally compatible with other international fora
5. Income-generating without explicit public sector funds
6. Acknowledge the demands of the developing world

[7] 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.

7. Stop being „DSI“ and become something else

Finally, she described the three proposed DSI scenarios stating that these scenarios were meant to present to the audience the spectrum of ideas that are compatible with open access and which meet the requirements of the 7 above stated project goals. Furthermore, in order to connect the WiLDSI project results to international discussions, the figures in the background paper were produced by the first global ABS Dialogue in Pretoria, South Africa (November 2019).[8]

Day 2

Dr. Jens Freitag gave a short recap of the previous day and then proceeded to explain the overview of Day 2 and how the group work would be divided. The primary objective of the group work was to test the three DSI open access scenarios, examine feasibility, propose new adaptations, and refine accordingly.

Testing the scenarios

Discussion of the 3 scenarios

Five groups (7-8 participants /group) were formed based on the professional background of the participants i.e. academics, policy makers and industrial representatives.

Each group discussed all three scenarios and assessed:

1. The practical implications/ consequences of the scenario in their everyday work?
2. Possible amendments
3. Ranking the scenarios according to the respective groups preference

After approx. 105 minutes of intense discussion, each group reported to the plenary.

Outcome of the group discussions

Some groups addressed certain general points (non-scenario-specific) that would need to be considered:

- With respect to any multilateral funds, who will pay, for what and how much?
- How does the DSI debate related and link to sovereign rights over GR?
- How does any DSI system relate to the existing ABS systems? It is important to keep in mind the perspectives of other international fora.
- There exists mistrust amongst the provider countries and this should be considered. We need to think of a fair solution if we want to get everyone on board.

[8] http://www.absinitiative.info/fileadmin//media/Events/2019/68_November_2019__PretoriaSouth_Africa/Report-First-Global-DSI-Dialogue-SouthAfrica-201911.pdf

A) Nagoya plus (Dialogue Option 2)

- This scenario is linked to the internationally recognized certificates of compliance (IRCCs) system. At present this would be difficult and disadvantageous as the IRCC system is currently not working (few IRCC issuing countries, often absent or difficult to find legislation). The IRCC system would eventually need to be updated to respond to all the needs including an obligation to use one document that is machine-readable (enables programming logic to interpret legal conditions).
- This scenario would require tracking and tracing which is theoretically feasible but would require high costs and extraordinary data management.
- In terms of benefits, the system would likely outweigh the benefits.
- Being a traceable system, could provide a certain amount of legal certainty.
- Would increase jurisdiction shopping. In other words, users would start to filter out the data that is tagged in some way and only use "untagged" data.
- This system will cause huge delays unless rigid rules are set in place.
- An unresolved issue is how to track, measure and calculate the contribution of each nucleotide sequence to the commercial product?
- Would only be feasible if a highly standardized MAT would be in place for all NSD and if ABS conditions are universal or restricted to a defined set of IRCCs.
- No clear distinction between commercial and non-commercial use.
- One advantage of this scenario is that it builds off an existing system.
- This bilateral model could be attractive from a provider country's perspective as this model would connect benefits (if they accrue) to an individual country (rather than a shared fund).
- Part of the compliance burden falls on the researcher who uploads the data and has to make the link to the standardized IRCC.
- Users would need to read carefully the conditions and understand their obligations.
- Possible amendments:
    - To include a time limitation on obligations.
    - Parties could offer certain types of data use exemptions.

In all, it was agreed that this scenario was the least attractive of the three scenarios, and garnered the highest criticism.

B) Country tag (Dialogue option 4)

- Country tag increases the value of the data from a scientific perspective.
- A subscription fee could be considered a hindrance. Is it still considered open access if there is a link with a subscription fee?
- Difference between commercial and non-commercial use is not clearly defined.
- With respect to the funds:
    - Who pays and how much?
    - How to distribute the funds? How to dedicate the funds and to which countries?
    - Governance is unclear.
    - Developing countries would have to contribute to the funds too.
    - Are public research institutes expected to pay? Would it be possible for the government to pay some fee that then allows users in their jurisdiction to operate?

    - A big question for the companies, why would it be attractive to pay? How much to pay?
    - Is it possible to provide incentives for subscription for e.g. biodiversity conservation.
- Parties could misunderstand that country that uploads data (e.g. sequencing country) would benefit from the system. (Needs to be stated clearly that country of origin of the GR from where the DSI is generated).
- How much a country contributes to the system is not really a good measure of how benefits should be distributed?
    - Would this system incentivize data production (upload more data) rather than quality data?
    - Is it intended that how much data a country contributes would determine their benefits? If so, garbage/junk DSI might be encouraged?
- It is unclear what role the databases will play? How would this system work legally? How or would they police the terms and conditions?
- Possible amendments:
    - With respect to the BS issue, addition of certain notions (for e.g. change of intent at a certain stage in the process)
    - Address non-monetary benefits
    - One suggestion: to pilot such a system for a limited time period and see how it works.

C) De-coupled (Dialogue option 3)

- No tracking and tracing required. This system is very close to the philosophy of open access.
    - Preserves open access
- Comparatively cost-saving as no tracking and tracing involved (in terms of transaction costs)
- One could buy legal certainty depending on the actual modality of the system
- Voluntary nature of the payment
    - Politically this would be very challenging.
    - This option is attractive because of its simplicity and potential monetary benefit generation.
- Coalition of the willing
    - Make this commitment to the coalition of the willing stronger (e.g. Pledges)
    - Declaration/commitment of companies (prior to COP) that they will be willing to contribute to some sort of fund (for Biodiversity conservation)?
    - The idea of corporate responsibility of the companies. Companies might be happy to participate if built into the corporate responsibility programs. To introduce some sort of branding and recognition??
    - Include a broader and more diverse spectrum of users. Companies who make money using the services provided DSI. For e.g. Oil companies who use DSI to predict where to drill? These companies should also contribute towards a payment.
- How much money is needed?
- How much are people willing to pay for legal certainty?
- Think about guarantees and policing/enforcing of the mechanisms in terms of thinking of ways to keep people in the system.

- The chicken and the egg problem: Who will step in onto this first? Who will initiate?
- One way to make the voluntary system work is to leave the bilateral system in place and offer an opt-in option.
- Governance of fund not predefined. How will the money be distributed? Who will collect the money?
    - An international body that governs the funds??
    - One suggestion: Can the country tag in scenario B be a useful modality for the distribution of funds?
    - Use the funds for capacity building in countries that really need it.
- One downside is that money is shown late as compared to scenario B where money appears earlier (upfront payment via subscription versus at point of commercialization).
- Depending on the actual governance/modality of the mechanism could convince more countries to step into this proposal. Make it more specific.
- How to measure the contribution of specific DSI to the final commercialized product?

Summarizing, there was a strong and mutual agreement on preserving and maintaining open access. There was also a strong preference for a multilateral system as compared to the bilateral system given the problems of scale and defining trigger points discussed on the first day. It was also agreed that country tag should be included as it enhances the scientific value. There were intense discussions regarding the redistribution of monetary funds with respect to:

- Countries that upload high amounts of DSI
- Country of origin of GR of the DSI
- Developing countries regardless of their DSI contribution
- Bilaterally at point of commercialization.

It was generally agreed that money should be used for biodiversity conservation as well as capacity building and technology transfer.

In terms of ranking the proposed scenarios (in order of most preferred to least preferred):

Group I: B, C and A

Group II: B, with A and C at the same level.

Group III: preferred not to rank

Group IV: C, B and A

Group V: B, C and A

Scenario B was the preferred option although there was a general consensus that Scenario B showed several caveats and still required several amendments. There was an interesting proposal of a hybrid scenario B and C i.e. a hybrid system of the 2 multilateral options. To merge the strengths of both the proposed MLS options and develop one hybrid proposal.

Synthesis and Way forward: Where do we go from here?

The workshop was concluded following an interactive discussion session conducted by Dr. Amber H. Scholz and George Haringhuizen. Questions that were addressed during this session were

- What is important for this stakeholder community with respect to what any potential DSI scenario offers?
- What should the proposed hybrid scenario encompass?

The following points were summarized based on the feedback and opinions from the participants:

- No tracking and tracing
- Open Access guaranteed
- Fairness (contribution of provider countries), shows "energy" of system
- Legal certainty for sale (at what price)
- Coalition of the willing from **both**:
    - Countries (laws)
    - Users (commitments)
- Country tag (GPS reporting systems)
    - INSDC informs and fund distribution based on indicators such as novelty, re-use and geography
- Elimination of jurisdiction shopping
- Include broader spectrum of users
- Corporate responsibility (branding/labelling)
- Market access fee (e.g. if you use biodiversity data)
- Transparency measures
    - Self-declaration of coalition of the willing
    - Country tag
- Monetary options
    - Subscription
    - Certification scheme
    - Micro-levy
    - Option (but more expensive) to contribute at time of commercialization (but want to avoid free rider problem and encourage early financial returns accrue, too hard to regulate)
    - Trigger?
- Eliminate/decreased compliance.
- Obligatory MLS
    - Different sectors pay different rates

Summary

The workshop on “Digital Sequence Information, Open Access, and Sustainable Benefit Sharing: Scientific Input to International Policy Decisions” provided a platform to the European scientific community to discuss their views with respect to the political dilemma around the DSI issue as well as offer their expert opinion on the three DSI scenarios. Summarizing, there was a strong and mutual agreement on preserving and maintaining open access. There was also a strong preference for a multilateral system as compared to the bilateral system. In conclusion, with respect to the three proposed scenarios, participants put forth the notion of merging the different aspects from the two proposed multilateral scenarios and to develop and describe a hybrid scenario in further detail.

Glossary of acronyms

ABS	Access and Benefit Sharing
AHTEG	Ad Hoc Technical Expert Group
BBNJ	Biological Diversity of Areas Beyond National Jurisdiction
BOLD	Barcode of Life Data System
CBD	Convention on Biological Diversity
COP	Conference of Parties
DDBJ	DNA Data Bank of Japan
DSI	Digital Sequence Information
EBI	European Bioinformatics Institute
EMBL	European Molecular Biology Laboratory
ENA	European Nucleotide Archive
EVA	European Virus Archive
CGRFA	Commission on Genetic Resources for Food and Agriculture
GBF	Global Biodiversity Framework
GISAID	Global Initiative on Sharing All Influenza Data GR Genetic Resource
GR	Genetic Resource
HMOs	Human Milk Oligosaccharides
INSDC	International Nucleotide Sequence Database Collaboration
IPIN	Institut Pasteur International Network
IRCC	Internationally Recognized Certificate of Compliance
ITPGRFA	International Treaty on Plant Genetic Resources for Food and Agriculture
MAT	Mutually Agreed Terms
MLS	Multilateral System of Access and Benefit-sharing
MTA	Material Transfer Agreement
NCBI	National Center for Biotechnology Information NP Nagoya Protocol
NRC	National Reference Centers
NSD	Nucleotide Sequence Data
OWEG	Open-ended Working Group on the Post-2020 Global Biodiversity Framework
PIC	Prior Informed Consent
SARS-COV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SNP	Single Nucleotide Polymorphisms
WHO	World Health Organization
WIPO	World Intellectual Property Organization
UNCLOS	UN Convention on the Law of the Sea
WiLDSI	**Wi**ssensbasierte **L**ösungsansätze für **D**igitale **S**equenz**i**nformation (Scientific approaches for DSI)