Leopoldina
Nationale Akademie
der Wissenschaften

Ad hoc Statement | 2021

Maintaining open access to Digital Sequence Information

Multilateral benefit sharing and Open Science

www.leopoldina.org/en

Publishing information

Publisher:
Deutsche Akademie der Naturforscher Leopoldina e. V.
– German National Academy of Sciences –
Jägerberg 1, 06108 Halle (Saale), Germany

Editing:
Dr. Henning Steinicke, National Academy of Sciences Leopoldina
Dr. Stefanie Westermann, National Academy of Sciences Leopoldina

Translation and proofreading:
Peschel Communications GmbH, Freiburg

Typesetting and layout:
unicommunication.de, Berlin

ISBN: 978-3-8047-4220-8

Bibliographic information published by the German National Library:
The German National Library lists this publication in the German National Bibliography.
Detailed bibliographic data is available online at http://dnb.d-nb.de.

Recommended citation:
German National Academy of Sciences Leopoldina (2021): Maintaining open access to Digital Sequence Information – multilateral benefit sharing and Open Science. Halle (Saale)

Contents

Summary and recommendations 4

1. Introduction 6

2. The Convention on Biological Diversity and the role of Open Science 8

3. The role of globally interconnected DSI databases 11

4. Background information: The importance of Digital Sequence Information 14

4.1. The importance of Digital Sequence Information for biodiversity research and conservation 14
4.2. The importance of Digital Sequence Information for natural compound research 15

5. Regulatory models for the use of Digital Sequence Information 17

5.1. General user fees 18
5.2. Selective user fees 19
5.3. Fees for research projects and work materials 19
5.4. Introduction of regulatory barriers 20
5.5. Conflicts of objectives with respect to the Convention on Biological Diversity and the 2030 Agenda 21

6. Conclusions 23

Members of the working group and other contributors 25

References 26

Summary and recommendations

Generating knowledge through publicly financed fundamental research is a pillar of modern society. This knowledge, which is applied in ways which are not always immediately apparent and whose results cannot be directly monetised, nevertheless forms the basis of a significant proportion of our prosperity. Knowledge generation, particularly in the life sciences, relies heavily on data and information relevant to research being openly accessible wherever possible (Open Science with Open Access and Open Data). In the majority of life science disciplines, open access to molecular information, and especially to globally available Digital Sequence Information (DSI) and Nucleotide Sequence Data (NSD)[1], is essential. This is because knowledge is often only produced following the analysis and comparison of a large amount of sequence information. This is often the only way of identifying new species of living organisms and of tracking changes in ecosystems.

Like the principle of Open Science, the conservation and sustainable use of biological diversity are highly accepted and supported around the world, and they represent important objectives of the *Convention on Biological Diversity* (CBD) and the *Nagoya Protocol*, which has now been ratified by most countries worldwide. One of the three pillars of the *Convention on Biological Diversity* is the equitable sharing of the benefits arising from the use of biodiversity. The *Nagoya Protocol*, which is binding under international law for the signatory states, provides the framework for this benefit sharing. As part of the process of developing an effective monetary benefit sharing, a restriction to open access to Digital Sequence Information (DSI) is currently under consideration. This conflicts with the international consensus described above. Open access to DSI must be maintained, not least in the interest of conserving biodiversity. Barriers to access would not only hinder the UN's Open Science objective, but would also profoundly impede biodiversity research due to the unwanted control effects of such restrictions.

The Leopoldina expressly acknowledges the equitable benefit sharing between all stakeholders – open access to Digital Sequence Information is an important non-monetary form of this. The global fund for conserving biodiversity currently being discussed by the contracting parties to the *Convention on Biological Diversity* is also welcomed without reservation. If companies generate income from biological diversity, a proportion of the profits should be paid into this fund. It is vital, however, that this fund is not built up establishing any kind of fee for accessing Digital Sequence Information. Instead, given the importance of multilateral conservation efforts, the contracting parties – and particularly those in the Global North – should demonstrate that they accept their responsibility for conserving biodiversity by contributing public money to the fund.

---

1 The term "Digital Sequence Information" and its abbreviation "DSI" are used in this statement in accordance with the meaning of the term as established in the context of the *Convention on Biological Diversity*. Further information on the correct scientific use of the terms can be found at: https://www.vbio.de/fileadmin/user_upload/wissenschaft/pdf/190601_DNFS-VBIO-Leibniz_LVB_DSI_Submission_of_views.pdf

Information on the place of origin is currently missing for almost half of all Digital Sequence Information. To further develop the *Nagoya Protocol*, it is therefore crucial to implement mechanisms into sequence databases which make it possible to track subsequent changes to the intended use of sequences and to assign actual monetary benefits (such as those generated through patenting and commercial use) to countries more reliably than today, therefore making it easier to share the benefits. The *Global Initiative on Sharing All Influenza Data* (GISAID), which is very important in the field of epidemiology, is an example of how the scientists involved can also be documented alongside the place of origin. However, it is not possible to apply this approach more broadly because it currently does not permit the data sets to be freely transferred into open databases, which would then allow the data to be analysed in a different context. Nevertheless, steps need to be taken to ensure that monetary gain from using open databases can be tracked in the future. Solutions for this need to be developed by the scientific community.

The Leopoldina recommends that the following fundamental points be taken into account when negotiating the development of the *Nagoya Protocol*:

1. The principle of Open Science should be used as a guide for the development of international research infrastructure.
2. There is a conflict of interest between Open Science and benefit sharing through the regulation of access to Digital Sequence Information.
3. Open access to Digital Sequence Information should be maintained for researchers and should be more widely recognised as a non-monetary means of global benefit sharing.
4. The scientific community should look for solutions for implementing mechanisms into DSI databases which facilitate the identification of the place of origin and stakeholders.

International benefit sharing must take place without threatening either the conservation of biodiversity or Open Science and thus must not hamper the achievement of the global Sustainable Development Goals established by the United Nations in 2015 as part of the *2030 Agenda*.

1. Introduction

The *Convention on Biological Diversity* (CBD) was adopted in 1992. It has three main objectives: 1) the conservation of biological diversity, 2) the sustainable use of biological diversity and 3) the equitable sharing of the benefits arising from the use of biological diversity (see Chapter 2). The *Nagoya Protocol*, a framework for benefit sharing which is binding under international law, was passed in 2010. This framework applies to both the non-commercial and commercial use of genetic resources and thus also covers fundamental research. One main objective is to prevent genetic resources from being used without the consent of the country of origin or without the resulting benefits being shared. Although the regulations are a step in the right direction, they have thus far rarely delivered the intended equitable sharing of benefits. At the same time, differences in national regulations, unclear competences of national authorities and complex negotiation processes often hinder fundamental academic research.

The current regulatory approaches for benefit sharing laid down in the *Nagoya Protocol* are limited. Firstly, the original hope that direct monetary benefits could be derived from a single or a combination of a few genetic resources has proven to be largely unrealistic; this form of value creation has remained limited to a few specific cases. Secondly, the implementation of the *Nagoya Protocol* at the national level often conflicts with the concept of global research networks and the increasing digitalisation of life sciences research. Digital Sequence Information (DSI) is now established as the standard tool for fundamental research in the life sciences. DSI is the digital information collected on the order of nucleotides in the DNA of a specific organism. This data is stored in globally interconnected open databases, allowing genetic information to be studied around the world. As a result, genetic resources often no longer have to cross national borders to be used for research purposes. Another important point is that the USA – an important location for research and an economic powerhouse – has not ratified the *Convention on Biological Diversity* and is therefore also not subject to the *Nagoya Protocol*. Although a contracting country can require US scientists to comply with the provisions of the used as a guide *Nagoya Protocol* as a prerequisite to granting them access to its genetic resources, there is no such obligation for research conducted in the USA.

The parties to both the CBD and the contractually separate *Nagoya Protocol* are actively looking for ways of implementing benefit sharing more effectively than to date. Their attention has now turned towards Digital Sequence Information.[2] Some of the solutions proposed during the ongoing negotiations would jeopardise the current global open access to this data. The establishment of new financial, legal or technical barriers would be a huge step backwards for the globally connected scientific community. Not only would it adversely affect the life sciences in all countries worldwide, but it would also be contrary to the principles of Open Science. Barriers of this kind would endanger the conservation of biological diversity and impede the achievement of the global Sustainable Development Goals established as part of the *2030 Agenda*.

2 AHTEG (2020a).

2. The Convention on Biological Diversity and the role of Open Science

When the *Convention on Biological Diversity* (CBD) was adopted in Rio de Janeiro in 1992, the primary motivation was to stop the global loss of biological diversity. The *Convention on Biological Diversity* was a milestone, extending beyond the traditional approaches previously taken to conserve biological diversity. It set out three equally important main objectives: 1) the conservation of biological diversity, 2) the sustainable use of biological diversity and 3) the equitable sharing of the benefits arising from the use of biological diversity.[3]

In the CBD, the conservation of biological diversity was described for the first time as a common concern of humankind. The CBD also recognised that the sustainable use of biological diversity was a means of conservation and acknowledged the rights and practices of indigenous and local communities whose lifestyles depend on biological diversity and who have shaped biological diversity through their hunting, livestock breeding and cultivation practices, in some cases over thousands of years. It was only through giving these three objectives equal status that it became possible to ensure that a large number of states signed up to support biodiversity conservation efforts and that access to genetic resources could be secured. Equitable benefit sharing plays a key role in this. When the benefits arising from the use of these resources are shared equitably between countries, the motivation of stakeholders to protect this diversity extends beyond local use and valorisation as well as the inherent value of biodiversity.

With the *Nagoya Protocol,* the contracting parties adopted a binding framework for benefit sharing in 2010.[4] This framework applies to both non-commercial fundamental research as well as to bioprospecting, i.e. the search for and the commercial use of genetic resources. The *Nagoya Protocol* also aims to prevent biopiracy, i.e. the use of genetic resources without the authorisation of the country of origin or without benefit sharing.[5] In accordance with the *Nagoya Protocol*, access to genetic resources is now regulated on a bilateral basis in most countries worldwide.

3 UN (1992).

4 The *Nagoya Protocol* is a separate supplementary agreement to the *Convention on Biological Diversity*. CBD (2011).

5 Prominent examples of this include cases involving plants grown for agricultural purposes, such as basmati rice and aubergines: Jamil (1998); Abdelgawad (2012). Plants grown for agricultural purposes are, however, covered by their own international agreement, the *International Treaty on Plant Genetic Resources for Food and Agriculture* (ITPGRFA).

In addition to the three aforementioned objectives, the *Convention on Biological Diversity* covers other important matters, in particular the promotion of open, collaborative science and internationally equal technological and scientific structures.[6] The *Nagoya Protocol* also recognises the important contribution technology transfer and scientific cooperation make to sustainable development.[7] Furthermore, scientific knowledge is crucial to achieving the objectives of the *Convention on Biological Diversity*. Accordingly, relevant gaps in knowledge and the resulting research needs have been identified for a wide range of topics at each of the regular *Conventions of Parties* (COP).[8]

Within the global scientific community, the resources in each individual country vary widely, not least due to differences in the amount of public research funding available. Nevertheless, science has developed enormously over the last few decades. Current projects studying biodiversity in countries with a high species richness are a particular example of this, as they largely involve equal cooperation between research groups in each of the participating countries, not least due to the CBD and the *Nagoya Protocol*. At the same time, modern communication, the globalisation of methods and the Open Science movement have all paved the way for more equitable access to research data and information worldwide. Many financial, technical and legal barriers to access have been and are continuing to be broken down. Despite this, inequality still exists, but both the scientific community and the United Nations have set themselves the goal of further reducing it.[9]

6 UN (1992). The overriding objectives specified in the CBD are the general promotion of science (Art. 12 Research and Training, Art. 18 Technical & Scientific Cooperation) and, above all, the free and unhindered access to data and information (Art. 17 Exchange of Information: "The Contracting Parties shall facilitate the exchange of information, from all publicly available sources, relevant to the conservation and sustainable use of biological diversity […]"). The role played by knowledge transfer is also highlighted in the biological diversity targets set in 2010 (Aichi Target 19).

7 CBD (2011): "The Parties to this Protocol, […] Recognizing the important contribution to sustainable development made by technology transfer and cooperation to build research and innovation capacities for adding value to genetic resources in developing countries, in accordance with Articles 16 and 19 of the Convention […]."

8 For more information on this, see the results of a study undertaken by the Institute for Biodiversity – Network e. V. (ibn) on behalf of the German Federal Agency for Nature Conservation [Bundesamt für Naturschutz, BfN]: https://biodiv.de/biodiversitaet-infos/forschungsbedarf-der-cbd.html

9 UNESCO et al. (2020).

Open Science

The importance of Open Science is growing worldwide. The aim of the concept is to generate knowledge transparently and to make it freely accessible – for the benefit of both research and society. Open Science is based on the belief that scientific progress can be made more quickly and comprehensively if existing knowledge is shared and developed in collaborative networks. Its key pillars include open access to scientific publications (Open Access), research data (Open Data), scientific software (Open Source), methods (Open Methodology) and course content (Open Educational Resources) as well as the open quality assessment of scientific papers (Open Peer Review). Open means that there are no or very few financial, technical and legal barriers to the access, reuse and further development of scientific findings and data. The concept thus allows research data, processes and results to be reused, shared and reproduced. Open Science enables potentially anyone to comprehend scientific findings. This makes it easier to transfer knowledge to business and society (Open to Society). In terms of DSI databases, Open Access is provided when there are virtually no barriers to access in terms of costs or restrictions on data sharing.

The United Nations see the principle of Open Science as an effective instrument for reducing injustices and inequalities.[10] Furthermore, Open Science is a cornerstone for achieving the Sustainable Development Goals set by the United Nations as part of the *2030 Agenda*.[11]

10 Amann et al. (2019); UNESCO (2020).

11 UNESCO (2017); UNESCO (2020).

3. The role of globally interconnected DSI databases

Today's scientific community is globally interconnected, and scientific progress is largely dependent on this worldwide network. Researchers in all scientific disciplines rely on being able to acquire available knowledge rapidly and on making their own findings available to the scientific community as quickly as possible without any restrictions. Open access to knowledge therefore plays a key role in this process. This applies in particular to life sciences research. In the majority of life sciences disciplines, access to molecular information, and especially to globally available Digital Sequence Information (DSI) and Nucleotide Sequence Data (NSD), is essential. In these fields, knowledge is often only produced following the analysis and comparison of a large amount of sequence information.

Digital Sequence Information (DSI)

The genetic information in all living organisms (deoxyribonucleic acid, DNA) consists of individual building blocks, known as nucleotides, which contain one of four bases: A, C, G or T. These are responsible for encoding each organism's characteristics. Decoding this genetic information requires DNA sequencing, which is the process of determining the order of nucleotides in the DNA. The sequencing of genetic information is a gradual process, in which shorter sequences of nucleotides are decoded one after the other and eventually combined into an entire sequence (genome). The complete sequencing of even larger genomes can now be conducted relatively inexpensively and quickly thanks to high-throughput methods.

The analysed sequences must be transferred into the relevant databases and made available by the time the scientific findings acquired with their help are published; otherwise, the manuscripts will not be accepted by scientific journals. As part of this process, publicly funded databases guarantee that research results remain permanently available and thus verifiable. The *International Nucleotide Sequence Database Collaboration* (INSDC) is a consortium which constitutes the central global infrastructure for the storage of sequence information. It comprises the three largest databases: GenBank in the USA, ENA in Europe and DDBJ in Japan. Analyses of all 743 currently known nucleotide databases have shown that 95 percent have shared their data directly through the INSDC; the remaining databases use sequences directly recorded in the INSDC and are therefore also dependent on this central infrastructure. The funds required by the service provided by the INSDC, which is openly accessible to scientists worldwide, currently amount to approximately 50 to 60 million US dollars a year and are provided exclusively by the USA (GenBank), Japan (DDBJ) and the member states of the *European Molecular Biology Laboratory* (EMBL ENA). At present, the inventory of sequences holds more than 1.5 billion sequences and is accessed by 10 to 15 million users a year.[12] Metadata, such as information on the origin of the data, must also be provided when recording sequences in a database.

12 The number of users was estimated using uniquely identifiable IP addresses. See Scholz et al. (2020); AHTEG (2020b).

These days, DNA can be sequenced relatively quickly and inexpensively. The resulting sequences are then entered into one or more of the available databases so that they can be accessed by the global scientific community (see box: "Digital Sequence Information"). Most databases are openly and easily accessible. They provide free, anonymous access and allow the sequence information they contain to be transferred to other databases. Only a few databases, such as the *Global Initiative on Sharing All Influenza Data* (GISAID), have regulatory barriers and require users to register, which means that they cannot be directly interconnected with open databases (see box: "Global Initiative on Sharing All Influenza Data"). However, this interconnection is essential for the INSDC and the many databases linked to it. It allows large sections or even the complete database to be mirrored, offering protection against data loss and ensuring faster access to the data sets. The interconnection also enables bioinformatics research into innovative approaches to data management and analysis. There are no financial barriers to access because the money needed to operate and curate the INSDC databases and to run the associated infrastructure is provided in the form of public funding from the countries involved (see box: "Digital Sequence Information").

Nowadays, the databases contain data from every country worldwide.[13] It is the open accessibility of this data to the scientific community and, above all, the opportunity this offers to compare it with other sequence information which makes it valuable.

Studies show that the patterns of data provision and data retrieval contradict the assumption that the genetic resources come predominantly from biodiverse countries, with the majority of users based in industrialised countries. On the contrary, more than 50 percent of the sequence information with a known country of origin comes from the USA, China, Japan or Canada and a little more than 50 percent of users are based in countries which do not finance the databases.[14]

13 Scholz et al. (2020).

14 Scholz et al. (2020).

Global Initiative on Sharing All Influenza Data (GISAID)

When the avian influenza virus A-H5N1 threatened to become a pandemic in 2005, researchers were confronted with restricted public access to the new genetic sequences analysed for this virus strain. This was because scientists in the countries where cases of avian influenza first originated were reluctant to share the latest information through openly accessible public domain archives such as ENA, DDBJ and GenBank. This was due to the fact that in these and other comparable sequence databases, the data is used anonymously and its usage could not be traced very easily, if at all. Researchers therefore believed that it was almost impossible to guarantee that the data suppliers' rights and interests would be effectively protected by the existing infrastructure. This meant that there were no sufficient incentives to encourage data sharing. Given the global spread of avian influenza, a new approach was urgently needed to overcome these obstacles. In response to this, the *Global Initiative on Sharing All Influenza Data* (GISAID)[15] was established in 2008.

Like other public domain archives, GISAID provides access to its data free of charge. However, in contrast to other sequence databases, this database is only available to users who identify themselves and agree to the GISAID sharing mechanism. This protects the data suppliers' intellectual property rights to the sequences they share and requires the data users to name the laboratory which originally made the specimens available and generated the sequence data and metadata. In addition to citing the relevant laboratories in their scientific publications, researchers are obliged to mention in their acknowledgements all research groups involved in making the data available and are also required to make every effort to include them in the scientific exploration of the sequence data. Moreover, all users of the database must consent to the fact that access to the data made available in GISAID is not subject to any restrictions. This is in the interests of promoting collaboration between researchers on the basis of the open sharing of data within GISAID as well as in respect of the rights and interests of all stakeholders. The aim of this is to ensure that the results derived from the data are used fairly for scientific purposes. The GISAID Database Access Agreement does not include any provisions on monetary compensation for the commercial use of sequence data.[16] At the same time, this type of benefit sharing is possible because both the geographical origin of the genetic resources and all the stakeholders can be identified.

The drawback of the concept is that the restrictions laid down in the terms of use prevent the data from being linked directly with open databases. This means that it is not possible for the data to be integrated and analysed more broadly in other contexts. This prevents the GISAID concept from being applied on a general level to data relevant to biological diversity, which is substantially more heterogeneous than the data sets stored in GISAID.

For several years, the *World Health Organization* (WHO) has been using sequences in GISAID to select which virus strains will be used to develop influenza vaccinations every six months. Researchers studying the coronavirus (SARS-CoV-2) also turn to the GISAID database as it holds the world's leading collection of sequences belonging to this virus type.

15 https://www.gisaid.org

16 https://www.gisaid.org/registration/terms-of-use/

4. Background information: The importance of Digital Sequence Information

Data analysis is becoming increasingly important in the life sciences, and bioinformatics has already become an essential part of these disciplines of study. High-throughput methods, known informally as the omics, have made a huge contribution to this. They allow certain information to be captured in full, such as the complete genome of a living organism (genomics) or all proteins produced in a cell type or a complete organism at a certain point in time (proteomics). The emergence of these methods led to an abundance of primary data being produced in a short amount of time, with new data being generated at a rapid rate. The data collected in this way must be processed using bioinformatics in order for the information it contains to be usable. These days, gaining new knowledge in the life sciences relies heavily on the evaluation of digital data sets by comparing them with other data sets.

Against this backdrop, sequence databases have become crucial in the life sciences over the past few decades. On its own, an organism's sequence data is just information without any appreciable scientific or economic value. It is only once this data is functionally classified and compared with other sequence information that it becomes scientifically valuable for fundamental research purposes. This is why scientists endeavour to enter the sequences they determine and the associated metadata into a sequence database as quickly as possible.[17]

4.1. The importance of Digital Sequence Information for biodiversity research and conservation

Knowledge of which species exist in the first place is an important prerequisite for the effective conservation of biodiversity. Taxonomy – the science of identifying, naming and classifying living organisms – is a crucial part of this,[18] and is itself increasingly reliant on the analysis of genetic data in the form of Digital Sequence Information. With the help of DSI databases, the sequences of a specific organism can be compared with the sequences of thousands of other species in order to ascertain whether a previously undiscovered species is present.

In addition to the identification and description of species, the phylogenetic classification of how they relate to other species is vital to our understanding of biological diversity and functional mechanisms within ecosystems. This also requires access to DSI databases.

17 Recent studies show that scientists compare an average of 44 sequences from different places of origin when performing a database query. Scholz et al. (2020).

18 See Leopoldina (2014) for detailed information on this.

Fundamental ecological research also constitutes a part of biodiversity research. For example, digital information can be used to investigate processes at a cellular or molecular level, which mirror functions within and the functioning of ecosystems. Modern high-throughput methods also allow all genomes within a given ecosystem to be identified (metagenomics), enabling quick insights into the ecosystem's current biological diversity (e.g. soils, bodies of water). Again, this type of data only becomes scientifically valuable once it can be compared with other data.

Sequence information can be translated into DNA barcodes. This allows the known species in a specific ecosystem to be determined quickly, which is crucial for monitoring, detecting and studying changes in ecosystems. Conservation research, which focuses on developing and testing effective forms of biodiversity protection, also uses Digital Sequence Information for various purposes, including for identifying fragmentation effects in ecosystems and for investigating the effective size of protected areas.

In all areas of biodiversity research, the information only becomes valuable once it is possible to compare this data with other data.

4.2. The importance of Digital Sequence Information for natural compound research

The investigation of natural and active substances plays a particular role in the context of the *Convention on Biological Diversity* and the *Nagoya Protocol* because it offers great potential for value creation from genetic resources. Plants, fungi and microorganisms (as well as some animals such as cone snails and scale insects) are especially relevant to researchers of natural and active substances. In addition to examining substances relevant to medicine,[19] natural compound research explores substances used in aroma, flavour and plant protection.[20]

The importance of DSI databases to active substance research can be clearly illustrated by the example of microorganisms. Microorganisms are able to produce a vast number of organic small molecules. These compounds are natural substances or secondary metabolites which form the basis of many important medicines, including antibiotics and cytotoxic agents. The genes for synthesising these compounds are arranged in clusters in the genome. The sequencing of thousands of genomes of microorganisms led to the discovery that many of these can encode between 20 and over 100 of these gene clusters for the biosynthesis of various secondary metabolites in their genome. This means they have the potential to form numerous and, in some cases, undiscovered substances.

19 Artemisinin and curcumin are examples of active substances relevant to medicine. Artemisinin is derived from the sweet wormwood plant, which is grown in China, Vietnam and some countries in East Africa. Curcumin is an active substance from the turmeric plant and was already the subject of a patent dispute in the 1990s.

20 Further examples are cited by Houssen et al. (2020).

However, when microorganisms are cultivated in pure cultures in a laboratory, they often only produce a fraction of these compounds; the gene clusters remain inactive and therefore do not cause secondary metabolites to form. On many occasions, other organisms (ecological context) are needed to activate these otherwise dormant gene clusters. In the vast majority of cases, the ecological context required is still unknown. Taking into consideration the microbial genomes known to date and their encoded gene clusters, it can be assumed that microorganisms have the potential to form countless, as yet undiscovered compounds, which could be used as active substances. The availability of genome sequences in open databases allows them to be studied using gene technology, which in some cases already allowed to activate dormant gene clusters, thus enabling the identification and isolation of new substances.

Methods such as this are extremely important in antibiotic research since 55 percent of all antibacterial agents are based on natural substances, most of which are microorganisms.[21] At the same time, this example demonstrates the relevance of publicly funded research because private companies are rarely involved in the search for antibiotic agents or the development of new antibiotics presently, not least due to the stark decline in profit expectations in this line of work.[22]

The research on microorganisms further highlights a more fundamental problem. In accordance with the *Nagoya Protocol*, it is very important to record the origin and geographical distribution of genetic resources. However, microorganisms are generally not limited to certain localities or regions and instead can be spread widely by air and water. These weaker regional ties make benefit sharing difficult or even impossible. Although the provenance of an individual specimen can be traced easily, identical genome sequences can be found in places which are thousands of kilometres away from each other.

21 Moreover, 30 percent of all anticancer drugs are based on natural substances, see Newman & Cragg (2020).

22 Academy of Sciences and Humanities in Hamburg & Leopoldina (2013); Simpkin et al. (2017).

5. Regulatory models for the use of Digital Sequence Information

The fundamental objectives of the *Nagoya Protocol* are the control of access to genetic resources and the equitable sharing of benefits arising from their use. It calls for the development of a mechanism which allows benefits to be shared equitably without hindering the other overriding objectives of the *Convention on Biological Diversity*.

Digitalisation in the life sciences and improvements in synthetic biology have resulted in a loss of importance of traditional methods of biopiracy, for instance the illegal export of crops. Genetic information can be sequenced locally and imported into globally accessible open databases. This means that living organisms or their genetic source material no longer need to cross national borders. Theoretically, this information can be used to synthesise a sequence and its encoded proteins anywhere in the world, although it is not possible to synthesise a complete living organism.[23] However, this information only gains scientific value once it has been compared with other sequences. Considerations are being made as to whether access to sequence databases – on which a large number of scientists worldwide rely every day – should be used to generate financial resources for benefit sharing. If this avenue were chosen, solutions would need to be found which would not limit open access to Digital Sequence Information so as to avoid hindering the overriding objectives of the *Convention on Biological Diversity* and imposing extensive restrictions on life sciences research. At the same time, solutions should be found which are not based on using research funds for a monetary benefit sharing.

The worry is that making just some databases accessible for a fee, as is currently under discussion, would have significant ramifications. At present, scientists worldwide have access to a large number of open databases. This network of many small and several large databases works so well precisely because the databases are easily accessible and interconnected and they mirror each other to a large extent, i.e. they contain all or some of the data set found in other databases. This offers protection against data loss and ensures faster access to the data sets. This is also beneficial in the field of bioinformatics, as it allows for research into innovative approaches to data management and analysis. Introducing financial barriers and thus limiting Open Access would endanger the global interconnection of databases, the functionality of the entire system and the quality of life sciences research.

Moreover, placing databases behind paywalls would have a particularly negative impact on researchers in countries in the Global South, as quantitative studies show that more than 50 percent of users come from this part of the world.[24] Restricting access like this would also be highly detrimental to the efforts to build a trustworthy, trans-

23 Although whole genomes can be reproduced, it is not possible to reproduce complete organisms. Many of the cellular components needed for this, such as highly complex ribosomes, cannot be produced in laboratories.

24 Scholz et al. (2020).

parent scientific system because the Open Data approach is a key pillar of the Open Science culture. It would be much more difficult to publish scientific findings and replicate analyses because results can only be published in journals if the primary data has been entered into the relevant databases. If these databases were behind a paywall, this could be particularly problematic for scientists working in economically weaker countries.

The introduction of legal barriers would have similar repercussions. If access to Digital Sequence Information were regulated in line with the current bilateral implementation of the *Nagoya Protocol*, this would likely have serious implications, especially in the area of fundamental life sciences research. This is because accessing every single piece of sequence information would require the consent of the country of origin and its enforcement agencies. It is already clear that the provisions laid out in the *Nagoya Protocol* often cause significant delays to scientific projects, or even prevent them altogether. Areas of research such as modern biodiversity research, molecular genetics and evolutionary biology, which analyse large data sets of DNA sequences, would face insurmountable hurdles if they had to reach a bilateral agreement for every sequence. Another problem with bilateral solutions like this is that they would require sequences to be clearly allocated to a country of origin. This would be virtually inconceivable in the case of most microorganisms and marine organisms because, as described above, it is rarely possible to pinpoint these organisms to a specific geographical location (see Chapter 4.2).

5.1. General user fees

Various models for monetary benefit sharing arising from the use of Digital Sequence Information are under discussion. One option would be the introduction of user fees, which could be charged for either providing or retrieving data. While charging user fees to research institutes and groups in countries in the Global North would presumably be feasible without significantly restricting access, it is likely that an arrangement like this would be a much greater impediment to research in countries where science is less well funded.

For paywalls to work effectively, it must not be possible to circumvent them. The largest database accessible worldwide, GenBank, is operated by the USA, which has not joined the *Convention on Biological Diversity* or signed the *Nagoya Protocol*. Charging user fees for the two other large databases – DDBJ and ENA – would therefore likely cause them to lose prominence internationally by prompting researchers to use GenBank more often in the short or medium term. This would surely jeopardise the continuation of the INSDC network in its current form, breaking apart the entire global network of large and small databases.

Given the loss of importance of restricted, non-open DSI databases which could be expected, it is also questionable whether any appreciable funds would even be generated in this way. What's more, the additional administrative costs for managing the databases would need to be covered, possibly from a portion of the fees charged.

5.2. Selective user fees

Another option would be charging fees selectively for the use of sequences according to their geographical places of origin or imposing selective paywalls on data users from certain countries. This could prevent researchers from countries providing genetic resources from having to pay to use sequences from their own country. However, the practicalities of a solution like this would be very complicated because it would be necessary to ascertain the exact geographical origin of sequences and users. Of the 1.5 billion sequences currently available in the INSDC network, the geographical origin has only been determined for around 40 percent.

The distribution of the identifiable geographical places of origin of sequence information and users also shows that financial resources generated from selective fees would probably be more likely to be sent from the Global South to the Global North because more than 50 percent of the traceable sequences come from the USA, China, Japan and Canada. A further undesirable effect of selective user fees could also be that sequence information subject to a fee would be rejected by popular databases like GenBank, which would probably turn to measures like this in an effort to maintain open access to their data.

The administrative work involved in implementing selective user fees would constitute another significant problem. Firstly, database operators would have to factor in additional administrative costs, which would threaten the existence of smaller databases in particular. Secondly, scientists would have to deal with substantially more bureaucracy.

5.3. Fees for research projects and work materials

Another mode of generating financial resources under consideration is the micro-levy model.[25] It draws on models used in the protection of intellectual property rights and involves applying levies to work materials, such as sequencing machines and consumables, in order to generate financial resources. The benefit of this approach is that open access to the databases would still be maintained. However, this option has disadvantages as well. The methods, materials and equipment used for sequencing are constantly evolving and a regulatory model like this would need to take this speed of change into account to remain useful in the long term. A significant proportion of the financial

25 A more detailed description of this model is provided by Scholz et al. (2020).

resources generated would probably need to be set aside just to cover the associated administrative costs. The blanket nature of this approach would also cause problems, as fees would have to be charged for every piece of equipment regardless of its actual intended purpose. Moreover, institutions and companies which use Digital Sequence Information without generating any data themselves, or, in other words, without actually purchasing any work materials, would gain access to the data without incurring any additional costs. Conversely, institutions and companies which perform sequencing work without generating any Digital Sequence Information, such as those in the field of human genetics diagnostics, would also have to pay the levy.

Another option would be to charge fees to research projects which rely on Digital Sequence Information so that the money raised could be paid into a multilateral fund. Again, the benefit of this is that access to open sequence databases would not be affected. However, such fees would be detrimental to non-commercial research. Public research funds would ultimately be used to generate a monetary benefit in addition to the non-monetary benefit produced for the scientific community. Moreover, these fees would take funds away from the very scientific disciplines which make a significant contribution to the conservation of biological diversity. And finally, the global distribution of user numbers suggests that countries in the Global South would be likely to pay in more money if such a mechanism were in place than those in the Global North.

5.4. Introduction of regulatory barriers

The DSI databases in the INSDC network and the many databases linked to this network are openly accessible. Without this virtually barrier-free access to all databases, it would not be possible for them to be interconnected. This interconnected approach is essential because it not only facilitates the mirroring of complete data sets, or at least large sections thereof, but it also allows for more extensive data integration and analysis. This is particularly important in biodiversity research, which in some cases relies on very heterogeneous data. Some databases in very narrowly defined research fields can work very successfully despite having restricted access. However, these restrictions mean that they cannot be integrated into the aforementioned network (see box: “Global Initiative on Sharing All Influenza Data”).

Introducing barriers retrospectively to just one part of the global, openly structured database network, for example by making it mandatory for some users to register, would destroy the network’s integrity. For such an approach to work, all users across the entire network would have to be required to register. However, this would be an extremely complex undertaking, which would need to be supported by the whole scientific community.

Developments since 2010 demonstrate how regulatory matters can inadvertently have a detrimental effect on research. When implementing the *Nagoya Protocol*, some of the contracting parties have often developed complex and rather inefficient national procedures. The bodies responsible for regulating the *prior informed consent* (PIC) of the country providing genetic resources and the *mutually agreed terms* (MAT) vary from country to country and, in some cases, cause lengthy delays or even prevent research projects from being conducted.[26] In 2018, 67 renowned Brazilian scientists commented on the devastating consequences of bureaucracy on biodiversity research in their country in a publication written under the auspices of the Brazilian Academy of Sciences.[27] Comparable experiences from the field of fundamental life sciences research have been described in other countries.[28]

The examples highlight how complicated and bureaucratic rules often hamper scientific activities to an extent comparable to that of bans and can hinder the direction of research strategies in the medium to long term.

5.5. Conflicts of objectives with respect to the Convention on Biological Diversity and the 2030 Agenda

Limiting open access to Digital Sequence Information would not only have a huge direct impact on the life sciences, but would also conflict with the fundamental objectives of the *Convention on Biological Diversity*. The *Nagoya Protocol* recommends creating conditions which support research on biodiversity conservation. In particular, steps also need to be taken to support biodiversity research in developing countries. In this context, special mention is made of measures for simplifying access to biodiversity resources for non-commercial research purposes.[29] Any means of restricting scientists' access to DSI databases would, however, have the opposite effect and would severely obstruct the development of life sciences research, above all in countries in the Global South. Approaches like this would also contradict the principles of Open Science, especially Open Data. Open Science itself is a fundamental means of non-monetary benefit sharing, which helps all researchers worldwide and makes an important contribution to the achievement of the Sustainable Development Goals set by the global community as part of the *2030 Agenda*.[30] Charging fees to access Digital Sequence Information would interfere with the already widespread practice of global scientific collaboration, would make it difficult for researchers to access key open collaborative infrastructure and would have a serious detrimental effect on this infrastructure.

---

26 Unpublished studies from the Leibniz Research Alliance "Biodiversity" [Leibniz-Verbund Biodiversität] show that collaborative international projects experience delays of one year on average. Some projects never come to fruition.

27 Alves et al. (2018).

28 Pawar (2020).

29 CBD (2011).

30 UNESCO et al. (2020).

Due to their direct impact on biodiversity research, access restrictions would also directly conflict with the intentions of the *Convention on Biological Diversity*. If access were restricted, research activities would shift towards countries offering free access to their genetic resources, such as many countries in the European Union, or towards countries which have not joined the Convention or signed the *Nagoya Protocol*, with the USA being the primary example. Biodiversity research in the world's most species-rich countries would be progressively hampered, including for the local scientists. Efforts to conserve biodiversity would suffer a severe setback in these countries in particular as well as globally.

6. Conclusions

In view of the arguments presented above, the Leopoldina recommends that the following fundamental points be taken into account when discussing access to Digital Sequence Information and efforts to ensure equitable benefit sharing:

1. The principle of Open Science should be used as a guide for the development of international research infrastructure.
2. There is a conflict of interest between Open Science and benefit sharing through the regulation of access to Digital Sequence Information.
3. Open access to Digital Sequence Information should be maintained for researchers and should be more widely recognised as a non-monetary means of global benefit sharing.
4. The scientific community should look for solutions for implementing mechanisms into DSI databases which facilitate the identification of the place of origin and stakeholders.

A basic aspect of sustainable global development is the equitable sharing of benefits between the parties providing genetic resources and the parties using them. Within the scientific community, this mainly takes place on a non-monetary basis, and this open access to sequence information is an important part of benefit sharing. The cornerstones of this are equitable conditions for all stakeholders, cooperation on equal terms and the acknowledgement that generating data and making it available is a fundamental part of scientific work.

A central concern of the *Convention on Biological Diversity* and the *Nagoya Protocol* is the desire to strengthen Open Science as a key aspect of conserving biodiversity worldwide. The objective of Open Science is to provide the entire scientific community and society as a whole with unrestricted access to data and information (Open Access, Open Data, Open to Society). Given the importance of DSI databases to the life sciences, it is in the best interest of all countries that Open Access to Digital Sequence Information be maintained.

The approaches followed by the *Global Initiative on Sharing All Influenza Data* (GISAID) are an example of how certain scientific infrastructure can improve benefit sharing. GISAID enables both the stakeholders and the geographical countries of origin of the genetic resources to be identified. It is currently not possible to apply this approach broadly because it does not permit the data sets to be freely transferred into open databases, which would then allow the data to be analysed in a different context. Nevertheless, steps need to be taken to ensure that monetary gain from using open databases can be tracked in the future. Solutions for this need to be developed within the scientific community.

It is, however, crucial that a benefit-sharing mechanism not hinder scientific progress as a whole. In order to keep its system fully functional, even GISAID has excluded monetary benefit sharing from its model despite the fact that it stores sequences used to develop vaccinations and therefore creates value directly.

The global fund for conserving biodiversity currently being discussed by the contracting parties to the *Convention on Biological Diversity* is welcomed without reservation. If companies generate income from biological diversity, a proportion of the profits should be paid into this fund. It is vital, however, that this fund is not built up establishing any kind of fee for accessing Digital Sequence Information. Instead, given the importance of multilateral conservation efforts, the contracting parties – and particularly those in the Global North – should demonstrate that they accept their responsibility for conserving biodiversity by contributing public money to the fund.

The conservation of biodiversity is a task for humanity which relies on the principles of Open Science. Global methods of benefit sharing which restrict access to Digital Sequence Information and limit the amount of data available for conserving biodiversity are inadvisable. Barriers to free access to Digital Sequence Information, be they bureaucratic, technical or financial in nature, are detrimental to both our knowledge of biodiversity and our ability to conserve it.

Members of the working group and other contributors

Prof. Dr. Rudolf Amann ML	Max Planck Institute for Marine Microbiology, Bremen, Germany
Prof. Dr. Katrin Böhning-Gaese ML	Senckenberg Biodiversity and Climate Research Centre (SBiK-F), Frankfurt am Main and Goethe University Frankfurt, Germany
Prof. Dr. Axel Brakhage ML	Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute (Leibniz-HKI), Jena, Germany
Prof. Dr. Birgitta König-Ries	Institute of Computer Science, Friedrich Schiller University Jena, Germany
Prof. Dr. Jörg Overmann	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

Scientific contribution and editing

Dr. Henning Steinicke	National Academy of Sciences Leopoldina, Halle (Saale), Germany
Dr. Stefanie Westermann	National Academy of Sciences Leopoldina, Halle (Saale), Germany

Reviewers

Prof. Dr. Erwin Beck	Bayreuth Center of Ecology and Environmental Research (BayCEER), University of Bayreuth, Germany
Prof. Dr. Ralph Bock ML	Max Planck Institute of Molecular Plant Physiology, Potsdam, Germany
Dr. Christoph Häuser	Museum für Naturkunde, Leibniz Institute for Evolution and Biodiversity Science, Berlin, Germany
Prof. Dr. Thomas Lengauer ML	Max Planck Institute for Informatics, Saarbrücken, Germany
Dr. Axel Paulsch	Institute for Biodiversity – Network (ibn), Regensburg, Germany

Further advisors

The following people contributed important ideas to the work on this statement:

Dr. Jens Freitag	Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany
Dr. Amber Hartmann Scholz	Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany

The Leopoldina would like to thank everyone who contributed to this statement. We would particularly like to take the opportunity to thank the reviewers and advisors for their constructive comments.

ML = member of the Leopoldina

References

Abdelgawad, W. (2012). The Bt Brinjal case: The first legal action against Monsanto and its Indian collaborators for biopiracy. *Biotechnology Law Report*, *31*(2), 136 – 139. http://doi.org/10.1089/blr.2012.9926

AHTEG – Ad Hoc Technical Expert Group on Digital Sequence Information on Genetic Resources (2020a). *Report of the Ad Hoc Technical Expert Group on Digital Sequence Information on Genetic Resources*. https://www.cbd.int/doc/c/ba60/7272/3260b5e396821d42bc21035a/dsi-ahteg-2020-01-07-en.pdf

AHTEG – Ad Hoc Technical Expert Group on Digital Sequence Information on Genetic Resources (2020b). *Combined study on Digital Sequence Information in public and private databases and traceability*. https://www.cbd.int/doc/c/1f8f/d793/57cb114ca40cb6468f479584/dsi-ahteg-2020-01-04-en.pdf

Academy of Sciences and Humanities in Hamburg & German National Academy of Sciences Leopoldina (2013). *Antibiotics Research: Problems and Perspectives* [Statement]. Berlin/Boston. https://www.leopoldina.org/uploads/tx_leopublication/2013_06_17_Antibiotics_Research.pdf

Alves, R. J. V., Weksler, M., Oliveira, J. A., Buckup, P. A. et al. (2018). Brazilian legislation on genetic heritage harms Biodiversity Convention goals and threatens basic biology research and education. *Anais da Academia Brasileira de Ciências*, *90*(2), 1279 – 1284. https://doi.org/10.1590/0001-3765201820180460

Amann, R., Baichoo, S., Blencowe, B. J., Bork, P. et al. (2019). Towards unrestricted use of public genomic data. *Science*, *363*, 350 – 352. https://doi.org/10.1126/science.aaw1280

CBD – Secretariat of the Convention on Biological Diversity (2011). *Nagoya Protocol on access to genetic resources and the fair and equitable sharing of benefits arising from their utilization to the Convention on Biological Diversity*. https://www.cbd.int/abs/doc/protocol/nagoya-protocol-en.pdf

Houssen, W., Sara, R. & Jaspars, M. (2020). *Digital Sequence Information on genetic resources*: *Concept, scope and current use*. *Studies on Digital Sequence Information on genetic resources*. *Studies of the intersessional period 2019 – 2020*. https://www.cbd.int/abs/DSI-peer/Study1_concept_scope.pdf

Jamil, U. (1998). *Biopiracy*: *The patenting of Basmati by Ricetec* [Commission on Environmental, Economic and Social Policy – South Asia & Sustainable Development Policy Institute Working Paper Series 37]. www.jstor.org/stable/resrep00629

Leopoldina, National Academy of Sciences (2014). *Challenges and opportunities of integrative taxonomy for research and society:* Taxonomic research in the era of OMICS technologies [Statement]. Halle (Saale). https://www.leopoldina.org/uploads/tx_leopublication/2014_Stellungnahme_Taxonomie_EN_final_01.pdf

Newman, D. J. & Cragg, G. M. (2020). Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *Journal of Natural Product Research*, *83*, 770 – 803.

Pawar, P. (2020). Strict biodiversity laws prevent Indian scientists from sharing new microbes with the world. *Science*, 20.10.2020. https://doi.org/10.1126/science.abf3547

Scholz, A. H., Hillebrand, U., Freitag, J., Cancio, I. et al. (2020). *Finding compromise on ABS and DSI in the CBD*: *Requirements and policy ideas from a scientific perspective*, *WiLDSI Project*, *Full report*. https://www.dsmz.de/fileadmin/user_upload/Presse/WILDSI/Final_WiLDSI_White_Paper_Oct7_2020.pdf

Simpkin, V., Renwick, M., Kelly, R. & Mossialos, E. (2017). Incentivising innovation in antibiotic drug discovery and development: Progress, challenges and next steps. *The Journal of Antibiotics*, *70*, 1087 – 1096. https://doi.org/10.1038/ja.2017.124

UN – United Nations (1992). *Convention on Biological Diversity*. https://www.cbd.int/doc/legal/cbd-en.pdf

UNESCO – United Nations Educational, Scientific and Cultural Organization (2017). *Open access to scientific information*. http://www.unesco.org/new/en/communication-and-information/access-to-knowledge/open-access-to-scientific-information/

UNESCO – United Nations Educational, Scientific and Cultural Organization (2020). *First draft of the UNESCO Recommendation on Open Science*. https://unesdoc.unesco.org/ark:/48223/pf0000374409.page

UNESCO – United Nations Educational, Scientific and Cultural Organization, World Health Organization & United Nations High Commissioner for Human Rights (2020). *Joint appeal for Open Science*. https://en.unesco.org/sites/default/files/joint_appeal_for_open_sciences_fin_en_fin.pdf

Selected publications from the series on science-based policy advice

2020	**Ad hoc statements on the coronavirus pandemic (1 – 7)**
	Biodiversity and Management of Agricultural Landscapes – Wide-ranging action is now crucial
	Additive Fertigung – Entwicklungen, Möglichkeiten und Herausforderungen [Additive manufacturing – developments, opportunities, and challenges] (Only available in German)
	Centralized and decentralized components in the energy system. The right mix for ensuring a stable and sustainable supply
2019	**Towards a scientifically justified, differentiated regulation of genome edited plants in the EU**
	Über eine CO_2-Bepreisung zur Sektorenkopplung: Ein neues Marktdesign für die Energiewende (Only available in German)
	Climate targets 2030: Towards a sustainable reduction of CO_2 emissions
	Air Pollution and Health
	Reproductive medicine in Germany – towards an updated legal framework
	Warum sinken die CO_2-Emissionen in Deutschland nur langsam, obwohl die erneuerbaren Energien stark ausgebaut werden? [Why are CO_2 emissions only slowly sinking in Germany while renewable energies are being greatly expanded?] (Only available in German)
	What is the Significance of Nuclear Energy to the Future of Global Power Generation?
	Saubere Luft – Stickstoffoxide und Feinstaub in der Atemluft: Grundlagen und Empfehlungen [Clean air – nitrogen oxides and particulate matter in breathable air: Basic principles and recommendations] (Only available in German)
	Biomass: striking a balance between energy and climate policies. Strategies for sustainable bioenergy use
2018	**Governance for the European Energy Union**
	Privatheit in Zeiten der Digitalisierung [Privacy in the age of digitalisation] (Only available in German)
	Species decline in the agricultural landscape
	Joint Committee on the Handling of Security-Relevant Research: Progress Report
	Artificial Photosynthesis
2017	**Coupling the different energy sectors – options for the next phase of the energy transition**

All publications from the series are available for free in PDF format on the websites of the Leopoldina.

The German Academy of Natural Scientists Leopoldina e. V.
German National Academy of Sciences

Jägerberg 1
06108 Halle (Saale)
Phone: +49 (345) 472 39-600
Fax: +49 (345) 472 39-919
Email: leopoldina@leopoldina.org

Berlin office:
Reinhardtstraße 14
10117 Berlin

The Leopoldina originated in 1652 as a classical scholarly society and now has 1,600 members from almost all branches of science. In 2008, the Leopoldina was appointed as the German National Academy of Sciences and, in this capacity, was invested with two major objectives: representing the German scientific community internationally, and providing policymakers and the public with science-based advice.

The Leopoldina champions the freedom and appreciation of science. It promotes a scientifically enlightened society and the responsible application of scientific insight for the benefit of humankind and the natural world. In its interdisciplinary discourse, the Academy transcends thematic, political and cultural boundaries. It is also an advocate of human rights.

It is the role of the Leopoldina, in co-operation with other national and international organisations, to identify and analyse scientific issues of social importance. The Leopoldina presents its policy recommendations in a scientifically qualified, independent, transparent and prospective manner, ever mindful of the standards and consequences of science.

www.leopoldina.org

ISBN: 978-3-8047-4220-8