Elizabeth Karger

Study on the use of digital sequence information on genetic resources in Germany

in the project

Scientific and technical support on implementing the Nagoya Protocol - Part 1 “Digital sequence information and ABS”

UFOPLAN 2017 F&E-Vorhaben

(FKZ 3517810100)

On behalf of the
German Competent National Authority for the Nagoya Protocol
in collaboration with the
Institut für Biodiversität – Netzwerk e.V.

May 2018

ibn

BfN Federal Agency for Nature Conservation

Cover picture: **istockphoto.com/cherezoff**

Author's address:

Elizabeth Karger

Institut für Biodiversität – Netzwerk e.V.
Nussbergerstr. 6a
93059 Regensburg
E-Mail: info@biodiv.de
URL: www.biodiv.de

Scientific Supervisor:

Ass. iur. Ute Feit

Federal Agency for Nature Conservation
Division I 1.3 "Competent National Authority for the Nagoya Protocol"
Isle of Vilm
18581 Putbus
E-Mail: ute.feit@bfn.de

This publication is included in the literature database "DNL-online" (www.dnl-online.de).

Publisher: Bundesamt für Naturschutz (BfN)
Federal Agency for Nature Conservation
Konstantinstrasse 110
53179 Bonn, Germany
URL: http://www.bfn.de

Bonn, Germany 2018

Table of contents

List of Figures

List of Tables

List of Abbreviations

ABS	Access and Benefit Sharing
AHTEG	Ad-Hoc Technical Expert Group
BLAST	Basic Local Alignment Search Tool
CBD	Convention on Biological Diversity
COP 13	Thirteenth Conference of the Parties to the Convention on Biological Diversity
COP-MOP 2	Second Conference of the Parties serving as the Meeting of the Parties to the Nagoya Protocol
DDBJ	DNA Data Bank of Japan
DNA	Deoxyribonucleic acid
DSI	Digital sequence information on genetic resources
EMBL-EBI	European Bioinformatics Institute
EU	European Union
EU Regulations	European Union Regulation No 511/2014 and European Union Regulation 2015/1866
FAO	Food and Agriculture Organization
GMO	Genetically modified organism
GSD	Genetic sequence data
ITPGRFA	International Treaty on Plant Genetic Resources for Food and Agriculture
MAT	Mutually Agreed Terms
NCBI	National Centre for Biotechnology Information
Nagoya Protocol	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
OECD	Organisation for Economic Co-operation and Development
PIC	Prior Informed Consent
PIP Framework	Pandemic Influenza Preparedness Framework
RNA	Ribonucleic acid
SBSTTA	Subsidiary Body on Scientific, Technical and Technological Advice

WHO	World Health Organization

1. Introduction

1.1 Reason for the study

Rapid developments in science and sequencing technology in recent years have resulted in an increasing trend for genetic material to be sequenced and the resulting sequences to be used for research and development purposes. Given the availability of many sequences in publicly available databases, researchers can use sequences without ever requiring access to the physical genetic resources from which the sequences were originally obtained. This is what some refer to as the "dematerialisation" of research and development on genetic resources (Bagley and Rai, 2013; Mannheim, 2016).

The *Convention on Biological Diversity* (CBD) and the *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* (Nagoya Protocol) provide the international framework for the sharing of benefits arising from the utilization of genetic resources. The CBD and the Nagoya Protocol deal explicitly with access and benefit sharing (ABS) arrangements for "genetic resources", which could be interpreted as referring only to physical genetic material and not the digital sequences generated from that material (Spranger, 2017). This would mean that ABS obligations are not triggered under the Nagoya Protocol when these sequences are accessed, e.g. from public databases, and used for research and development. There is concern among some actors that the use of sequences could thus undermine the international ABS regime for genetic resources (Lawson and Rourke, 2016; Mannheim, 2016).

In recognition of the uncertainties surrounding the potential impacts that the use of sequences could have on the objectives of the CBD and the Nagoya Protocol, the Thirteenth Conference of the Parties to the CBD (COP 13) and the Second Conference of the Parties serving as the Meeting of the Parties to the Nagoya Protocol (COP-MOP 2) adopted parallel decisions in December 2016 to start an international process to explore this issue. These decisions refer to "digital sequence information on genetic resources", short "DSI". The acronym DSI has been used throughout this report but it should be kept in mind that DSI is more of a political term that has emerged from the CBD forum (Laird et al., 2018) and that the term is still under consideration at the international level.

DSI and its potential impact extend beyond the CBD and the Nagoya Protocol. For example, both the World Health Organization (WHO) and the Food and Agriculture Organization (FAO) are also investigating the potential impact of DSI (referred to as Genetic Sequence Data (GSD) and DSI) on their respective international instruments, namely the Pandemic Influenza Preparedness Framework (PIP Framework) and the International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA).

Germany is a party to both the CBD and the Nagoya Protocol. It does not regulate access to its own genetic resources but as a member of the European Union, it is required to implement the relevant European Union regulations[1] (EU Regulations) on compliance measures for users of genetic resources that fall under the scope of the Nagoya Protocol. Germany has also enacted its own national law[2] for this purpose.

Germany is typically regarded as a user country with respect to genetic resources, although genetic resources found on German territory are also used for research and development purposes. As a country with a strong research and development sector, it is very likely that actors in Germany who are involved in research and development on genetic resources also use DSI and would potentially be affected by the outcome of the international discussions on DSI.

1.2 Objectives of the study

This explorative study was conducted to generate some basic understanding of the German stakeholders who work with DSI. The main objectives were to:

- understand who the stakeholders in Germany are and how they work with DSI;
- understand the importance of DSI for their work;
- understand how DSI is accessed and shared by stakeholders in Germany; and
- determine whether any of these stakeholders have already had relevant experiences with DSI in the context of ABS arrangements under the Nagoya Protocol.

The study also attempted to shed some light on the choice of terminology, i.e. the term DSI, and its appropriateness as well as the response of stakeholders to the idea of ABS arrangements for DSI.

1.3 Methodological approach

This exploratory study was based on a review of the relevant scientific and grey literature and the responses of stakeholders to an online survey.

1 EU Regulation (EU) No 511/2014 of the European Parliament and of the Council of 16 April 2014 on compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilization in the Union and Commission Implementing Regulation (EU) 2015/1866 of 13 October 2015 laying down detailed rules for the implementation of Regulation (EU) No 511/2014 of the European Parliament and of the Council as regards the register of collections, monitoring user compliance and best practice

2 Gesetz zur Umsetzung der Verpflichtungen nach dem Nagoya-Protokoll, zur Durchführung der Verordnung (EU) Nr. 511/2014 und zur Änderung des Patentgesetzes sowie zur Änderung des Umweltauditgesetzes vom 25. November 2015

2. Background

2.1 The CBD and the Nagoya Protocol

The CBD, which entered into force in 1993, has three main objectives, namely the conservation of biological diversity, the sustainable use of its components and the fair and equitable sharing of the benefits arising from utilization of genetic resources (Art. 1 CBD). The Nagoya Protocol elaborates on the ABS system, which was established by the CBD.

Prior to the CBD coming into force, it was disputed whether international law treated biodiversity as the common heritage of mankind but the CBD clarified that States have sovereignty over their genetic resources (Art. 3 CBD). If parties choose to regulate access to their own genetic resources, they may require Prior Informed Consent (PIC) and Mutually Agreed Terms (MAT), which determine the relevant ABS arrangements (Art. 15 CBD). ABS was intended to provide incentives for the sustainable use and conservation of biodiversity by changing genetic resources from open access goods into private or club goods and creating a market for them (Richerzhagen, 2011). Sovereignty was intended to enable the allocation of any benefits arising from the use of genetic resources (De Jonge, 2013).

The CBD, however, did not provide detailed arrangements for implementation of the ABS system and in the years following its adoption, very little happened (Medaglia, 2015). In 2002, the non-binding Bonn Guidelines were adopted to provide some guidance on the implementation of ABS. In 2002, the World Summit on Sustainable Development called for a more detailed and legally binding ABS agreement. The negotiations on the binding instrument took six years and the Nagoya Protocol was finally adopted in 2010. It came into force on 12 October 2014 and at the time of writing, there were 105 parties to the Nagoya Protocol.

The Nagoya Protocol elaborates on the ABS system established by Art. 15 of the CBD. It has three main pillars, namely access to genetic resources, fair and equitable sharing of benefits arising from the utilization of those resources, and compliance. When the Nagoya Protocol was adopted, there were still some issues, e.g. regarding scope, that had not been fully resolved, which has meant that the implementation of the protocol has been somewhat controversial (Greiber et al. 2012). Although some argue that the Nagoya Protocol was intended to create trust between countries providing genetic resources and countries using genetic resources, others suggest that ABS in general has been a source of alienation between countries (Greiber et al. 2012; Rosendal and Andreson, 2016). Criticisms of the ABS system include the failure of parties to implement national legislation and to create relevant administrative procedures to enable users to access their genetic resources, the lack of compliance measures in user countries, and the high transaction costs associated with negotiating benefit sharing contracts (Greiber et al., 2012; Medaglia, 2015). Few successful cases of ABS have been officially reported and many expectations regarding benefit sharing, especially those relating to monetary benefits, have not been met (Richerzhagen, 2011).

It is against this background that the discussions on DSI are now taking place.

2.2 Definitions

2.2.1 Definition of "genetic resources"

The definitions in the CBD apply to the Nagoya Protocol (Art. 2 Nagoya Protocol). Art. 2 of the CBD defines "genetic material" as any material of plant, animal, microbial or other origin containing functional units of heredity. Genetic resources are a subset of genetic material (Glowka et al., 1994) and refer to "genetic material of actual or potential value" (Art. 2 CBD).

Human genetic resources do not fall within the scope of the CBD or the Nagoya Protocol. In accordance with Art. 4(4) of the Nagoya Protocol, specialised ABS Instruments prevail in respect of the specific genetic resource covered by the specialized instrument and for the purpose of that instrument. This currently includes plant genetic material for food and agriculture covered by the FAO's ITPGRFA and, at least according to the European Union, pandemic influenza viruses covered by the WHO's PIP Framework.

According to the CBD, "genetic material" contains functional units of heredity. Prior to the adoption of the Nagoya Protocol, there was a lot of debate about whether the term "genetic resources" referred only to deoxyribonucleic acid (DNA) or also to ribonucleic acid (RNA) and proteins, which do not contain functional units of heredity but retain information, as well as biochemical compounds that result from cellular metabolism (Morgera et al., 2014).

Although the CBD refers to the "utilization of genetic resources" and this term is relevant for determining the scope of ABS obligations, the CBD does not define it. A definition was included in the Nagoya Protocol. The "utilization of genetic resources" is defined in Art. 2 (c) as:

> *"to conduct **research and development** on the **genetic and/or biochemical composition of genetic resources,** including through **the application of biotechnology** as defined in Article 2 of the Convention". [emphasis added]*

The term "utilization of genetic resources" effectively expands the interpretation of genetic resources to also cover research and development on both the genetic and biochemical composition of genetic resources. It includes research and development involving biotechnology, which is defined in Art. 2(d) of the Nagoya Protocol as:

> *"any technological application that uses biological systems, living organisms, or **derivatives** thereof, to make or modify products or processes for specific use". [emphasis added]*

The definition of biotechnology includes the term "derivative", which is defined in Art. 2(e) of the Nagoya Protocol as:

> *"a naturally occurring **biochemical compound resulting from the genetic expression or metabolism of biological or genetic resources**, even if it **does not contain functional units of heredity**". [emphasis added]*

These definitions clarify that utilization of genetic resources is not necessarily restricted to research and development on material containing "functional units of heredity" (Greiber et al., 2012). Thus, the Nagoya Protocol widens the scope of "genetic resources" to include RNA and proteins as well as naturally occurring biochemical compounds that result from cellular metabolism, e.g. oils or resins (Morgera et al., 2014).

2.2.2 Definition of "research and development"

The term "research and development" is not defined by either the CBD or the Nagoya Protocol, which means that the ordinary meaning of these terms applies. A commonly cited definition is the Oxford Dictionary definition (Morgera et al., 2014), which is:

> *"the systematic investigation into and study of materials and sources in order to establish facts and reach new conclusions".*

The Frascati Manual from the Organisation for Economic Co-operation and Development (OECD) includes both basic and applied research in the definition of "research and development". Both definitions have been cited by the European Commission (European Commission, 2016) in order to further describe the meaning of research and development.

2.2.3 Definition of "digital sequence information on genetic resources"

"Digital sequence information on genetic resources" is the term used in the relevant COP/COP-MOP decisions from 2016 (COP13 XIII/16 and COP-MOP2 2/14). This term has not yet been defined and according to Laird et al. (2018) it is not used outside the CBD forum. As DSI has not been defined, it is not entirely clear what it refers to. There is recognition that there is uncertainty surrounding the terminology and it is being considered as part of the process put in place by the COP/COP-MOP decisions.

Databases around the world contain different types of sequences from wholly to partially sequenced genomes, DNA sequences, RNA sequences, protein sequences, etc. (Varshney et al., 2014). In addition to the sequences themselves, databases also have various amounts of metadata, annotations etc. which provide further information, e.g. about the relationship between genotypes and phenotypes, gene identity and function and mutations etc. Laird et al. (2018) suggest that the term DSI could include not only DNA sequences but also whole genome sequences, RNA sequences, exome sequences, degradome sequences and amino acid sequences as well as possibly including metagenomes, various epigenomic markers and other molecular information.

There are many different terms which are used to refer to sequences and many of these terms appear to be used interchangeably. Laird et al. (2018) point out that the term "data" is most widely used with respect to sequences, especially by people in the scientific

community. This raises the question as to whether it would be possible to make a distinction between data and information in the context of DSI. Manzella (2016) suggests that differentiation between data and information could be of relevance for the scope of any decision on the issue.

Whereas data may be defined as signs, patterns, characters or symbols which potentially represent some object or process, information can be defined as the meaning which is provided once data has been considered, either alone or in combination in a particular context (Baškarada and Koronios, 2013; Lawson and Rourke, 2016). The sequence of letters which represent the nucleotides can be regarded as being arbitrary but at the same time, it carries a genetic message that may ultimately transfer "meaning" to the person who receives the message (Yockey, 2005). According to Jaspars (2017), it may be possible to differentiate between various levels or categories of information relating to sequences depending on the amount of data processing or analysis that has taken place, e.g. ranging from raw data (e.g. DNA and RNA sequences and metadata) to data which has been analysed and annotated using an algorithm (e.g. protein structure or metabolite data) to data which have been analysed and critically evaluated by an expert.

There is a danger that the failure to clarify information concepts can potentially lead to legal uncertainty or instruments that cover inappropriate subject matter (Bygrave, 2015). Clarification of the scope of the term DSI, if indeed this term continues to be used, is necessary.

2.3 The decisions of COP13 and COP-MOP 2

Two parallel decisions were made regarding DSI at COP13 (Decision XIII/16) and COP-MOP2 (Decision 2/14) in Cancun, Mexico in December 2016. COP-MOP2 Decision 2/14 provides that the documentation generated according to COP13 Decision XIII/16 will be used to inform the deliberations of the upcoming COP-MOP3 about the potential implications of DSI for the Nagoya Protocol.

COP13 Decision XIII/16:

- invited parties, other governments, indigenous people and local communities, and other relevant organizations and stakeholders to make submissions on DSI;
- requested the Executive Secretary to compile and synthesize the submitted views;
- requested the Executive Secretary to commission a fact-finding and scoping study;
- mandated the creation of an Ad-Hoc Technical Expert Group (AHTEG) to:
  - consider the synthesis of the submitted views and the scoping study;
  - examine the potential implications of the use of DSI for the objectives of the CBD and the Nagoya Protocol;
  - consider the technical scope and legal and scientific implications of existing terminology related to DSI on genetic resources; and

- identify the different types of DSI on genetic resources that are relevant to the CBD and Nagoya Protocol.

The AHTEG will submit the outcomes of its discussions to the Subsidiary Body on Scientific, Technical and Technological Advice (SBSTTA) for consideration at its 22nd meeting in June 2018. COP13 and COP-MOP2 requested the SBSTTA to consider the outcomes of the AHTEG and to make recommendations on the potential implications of the use of DSI for the objectives of the CBD and the objective of the Nagoya Protocol. These recommendations will be considered at COP14/COP-MOP3, both of which are scheduled for November 2018.

All of the relevant documents generated through this process are made available on the CBD Secretariat's website (https://www.cbd.int/abs/dsi-gr/ahteg.shtml).

2.4 Emerging positions on DSI

While the terminology and implications of DSI for the objectives of the Nagoya Protocol are still under consideration at the technical level, conflicting views about whether the Nagoya Protocol covers DSI have already emerged.

Some parties have argued that DSI falls within the scope of the Nagoya Protocol. For example, during the discussions at COP13/COP-MOP2, Costa Rica stated that DSI relates to ABS and is covered by the Nagoya Protocol. The Philippines proposed that DSI relates to ABS but further clarification is needed as to how. Other parties argued that DSI falls outside the scope of the Nagoya Protocol, e.g. the European Union, Japan and New Zealand (IISD Reporting, 2016). Some of the arguments raised during the discussion included that of equivalence between genetic resources and DSI.

Further indications about the position of relative actors can be gleaned from the submissions made to the Secretariat, although some parties have not yet clearly defined their position and many parties did not make a submission at all.

Pending any relevant agreement at the international level, the European Union has advised users of genetic resources within the European Union that access to and utilization of DSI obtained from publicly accessible databases falls outside the scope of the Nagoya Protocol and the corresponding EU Regulations, if there was no access to or utilization of the genetic material from which the DSI was derived (European Commission, 2016).

2.5 Relevant sectors and uses

2.5.1 Relevant stakeholders

With a strong research base and various industry sectors, Germany is typically regarded as a user country with respect to genetic resources. The main stakeholder groups which use

genetic resources are basic researchers, collections of genetic resources and industry. In the process of developing guidance documents for the implementation of and compliance with the Nagoya Protocol and the corresponding EU Regulation, the European Commission identified seven relevant industry sectors for the use of genetic resources. These include animal breeding, plant breeding, biotechnology, biostimulants and biocontrol, cosmetics, food and feed, and pharmaceuticals. As actors in these stakeholder groups use genetic resources in research and development, it can be assumed that they may also access and use DSI in support of that work.

The various stakeholder groups are often characterised by high levels of complexity and they are not necessarily well-defined, meaning that actors may, for example, belong to more than one stakeholder group (European Commission, 2016). Especially the lines between commercial and basic research are not always clearly defined. Although actors might be treated as different stakeholders, the distinction between commercial and non-commercial research can blur on many different levels, e.g. at an institutional level but also at the level of the individual researcher (von Kries and Winter, 2015).

2.5.2 Relevant uses

It has been possible to sequence genetic material since the 1970s but more recent improvements in technology, such as high-throughput sequencing and third generation sequencing, allow longer sequence reads to be generated faster and at lower costs (Varshney et al., 2014). There are petabytes of data stored in databases around the world and the amount of data is increasing rapidly, transforming biology into a big data field (Marx, 2013). Disciplines such as bioinformatics support the life sciences and make work with big data possible (Ramsden, 2015).

Sequences are used in a wide range of research disciplines but given the complexity of these various disciplines, it is not possible to provide a detailed description of all possible uses here. For that reason, only a brief overview of some common uses of DSI has been provided.

DSI is used in taxonomic work. In recent years, many species have been sequenced with the aim of creating DNA barcodes, which are short DNA sequences from a standardized region of the genome. These DNA barcodes are compared to reference sequences to identify known species and potentially discover new ones (Zimmermann et al., 2014).

Phylogeny is the investigation of the evolutionary relatedness of groups of organisms and their common ancestors (Charleston, 2013). Whereas scientists traditionally relied on morphological, physiological and behavioral characteristics as well as paleontological records for phylogenetic reconstructions and classifications, DNA, RNA and protein sequences are now also used for this purpose (Brocchieri, 2001). In the study of evolution, sequences are used to investigate the concept of the species, how species are formed, how genes change and evolve over time as well as how species adapt to biotic and abiotic aspects of their environments (Tyler-Smith et al., 2015).

In population genetics, sequencing is used, for example, to examine variation among the members of a population and how this variation is maintained from one generation to the next (Amorim, 2013). Population genetics can potentially also aid conservation, allowing questions to be addressed such as effect size and the distribution of loci affecting fitness (Allendorf et al., 2010).

The new "omics" technologies allow thousands of datasets to be mined parallel and to incorporate different molecular activities, technological platforms and model organisms (Huttenhower and Hofmann, 2010). Genomics includes the study of whole genomes, the transcriptome, epigenetics, and functional genomics (Kaur, 2013). Transcriptomics focuses on gene expression (McGettigan, 2013) and epigenetics is concerned with heritable changes in gene expression where the DNA sequence itself does not change (Kaur, 2013). Functional genomics deals with understanding gene function and determining the relationship between the organism's genome and its phenotype (Kaur, 2013). Proteomics is concerned with analysing the proteins in a sample of genetic material and involves high-throughput approaches together with physical and biochemical techniques (Coorrssen, 2013). Metabolomics is the systematic identification and quantitation of all metabolites in an organism or biological sample. Although metabolomics deals specifically with molecules, it is regarded as being important for determining gene function and for the annotation of sequences (Idle and Gonzalez, 2007).

Most microbes have not been or cannot be cultured in laboratories, meaning that sequencing provides a way of accessing many of these organisms (Culligan et al., 2014). Metagenomics, which involves sequencing an environmental sample, is also being used increasingly. The whole sample may be sequenced, assembled and annotated, or alternatively, particular genes, gene families or useful markers may be targeted (Allan, 2014). Metagenomics has various uses, e.g. the observation of soil bacterial diversity, investigating microbes in the marine environment and determining the impacts of anthropogenic change or natural environmental fluctuations on microbial communities (Shokralla et al., 2012). Metagenomics is also a useful tool for the discovery of novel genes and metabolic pathways (Culligan et al., 2014).

DSI is also essential to the emerging field of synthetic biology (Scott et al., 2015), which is concerned with gene expression, gene networks, metabolic engineering and genetic circuits, synthetic genes and synthetic networks (Oldham et al., 2012). The boundaries between synthetic biology, biotechnology, and metabolic engineering are not necessarily clear but synthetic biology is essentially concerned with writing new genetic information (Welch et al., 2017), although it is more than just cutting and pasting DNA sequences of characterized parts (de Lorenzo and Danchin, 2008). The applications of synthetic biology are diverse and include production of chemicals, biofuels, medicines, plastics, polymers and rubbers as well as plant feedstocks for microbe consumption (Laird and Wynberg, 2016; Koenig et al., 2015).

Increasingly, sequences are used by biotechnology and pharmaceutical companies as the sequences may hold information about potential drug candidates (Laird et al., 2018). DSI may also play an important role in feeding the growing human population and meeting expectations regarding food quality. Sequences are used in plant breeding, e.g. to map populations and breeding lines, to explore the relationship between genotype and phenotype with much greater resolution and to move plant research beyond the traditional focus on model species (Varshney et al., 2014).

2.6 Possible consequences

2.6.1 Possible consequences of maintaining the status quo

Open access to DNA, RNA and amino acid sequences was established as the norm in the 1980s when databases such as GenBank, the European Bioinformatics Institute (EMBL-EBI) and DNA Data Bank of Japan (DDBJ) were first being established (Lawson and Rourke, 2016). If DSI remains outside the scope of the Nagoya Protocol, the status quo would essentially be maintained and for people using DSI for research or development purposes within the European Union, no compliance obligations would be triggered under the EU Regulations.

There are concerns among some actors that use of DSI for research and development purposes could undermine the ABS system if there are no ABS obligations (Lawson and Rourke, 2016). Some actors may regard the use of DSI in the absence of benefit sharing obligations as a form of digital misappropriation of genetic resources (Bagley, 2015).

The Nagoya Protocol is a floor level agreement, meaning that parties can go above and beyond the provisions of the Nagoya Protocol in their own domestic legislation (Morgera et al., 2014). In the absence of any international agreement regarding ABS for DSI, countries may choose to take a unilateral approach to somehow regulating access to DSI (Bagley, 2015). There are already examples of countries implementing ABS legislation which includes the informational elements of genetic material, e.g. Brazil (Mannheim, 2016).

2.6.2 Possible consequences of DSI being treated as falling within the scope of the Nagoya Protocol

If DSI is treated as falling within the scope of the CBD and the Nagoya Protocol, this would potentially result in the application of ABS arrangements to DSI. It is not clear how this would be implemented at a national level, whether PIC and MAT obligations would become relevant for DSI and what the implications for compliance would be.

There have been many criticisms of the existing ABS system for genetic resources, including restrictive legislation that prevents access to genetic resources as well as the lack of adequate administrative procedures which enable users of genetic resources to get PIC and MAT (Greiber et al., 2012; Medaglia, 2015; Richerzhagen, 2011). It is not clear whether similar issues would be faced with respect to access to DSI, if it is regulated in some way. Due to the large number of highly distributed actors involved and the large volumes of data

accessed, shared and used around the world, elements of ABS such as the identification of potential beneficiaries and the allocation of benefits could become extremely difficult (Welch et al. 2017).

Bagley (2017) suggests that a multilateral benefit sharing mechanism as envisaged in Art. 10 of the Nagoya Protocol could be an alternative option for DSI. In the submissions made to the CBD Secretariat, this possibility was also raised by some actors, e.g. by Africa and Mexico (CBD Secretariat, 2018).

Art. 10 of the Nagoya Protocol provides:

> *"Parties **shall consider** the **need for and modalities** of a **global multilateral benefit-sharing mechanism** to address the fair and equitable sharing of benefits derived from the utilization of genetic resources and traditional knowledge associated with genetic resources that occur in **transboundary situations** or for which **it is not possible to grant or obtain prior informed consent**. The benefits shared by users of genetic resources and traditional knowledge associated with genetic resources through this mechanism shall be used to **support the conservation** of biological diversity and the **sustainable** use of its components **globally**." [emphasis added]*

When the Nagoya Protocol was agreed upon, the possibility of creating a multilateral benefit sharing mechanism was left open by Art. 10. The provision was not negotiated fully by the parties but was added to the proposed text of the Nagoya Protocol at the end of the negotiations to accommodate some of the unresolved issues between the parties (Greiber et al., 2012). There have been ongoing discussions at the international level about Art. 10.

There are two possible situations in which a multilateral benefit sharing mechanism may be considered, namely where there are transboundary genetic resources and situations where it would not be possible to obtain PIC and MAT for genetic resources. Bagley (2017) suggests that DSI may be comparable to these situations.

Art. 10 is purely procedural, requiring only that parties consider the need for a multilateral benefit sharing mechanism and the possible modalities of such as mechanism. It leaves the parties substantial freedom in designing any such mechanism (Morgera et al., 2014).

3. The survey

This exploratory study was based on a review of the relevant scientific and grey literature as well as the responses of relevant stakeholders to an anonymous online survey. A survey instrument was developed for this study (see Annex D), although some inspiration and guidance for the questions was drawn from published surveys on data sharing (Tenopir et al., 2011; Tenopir et al., 2015; Fecher et al., 2015; Schmidt et al., 2016). An anonymous survey was chosen as it was thought that it would encourage people to participate in the survey and answer questions openly without having any concerns regarding attribution of their information or views either to them personally or their organization.

To provide a broad perspective on the use of DSI and the relevant issues, respondents were sought from the three main stakeholder groups in Germany, i.e. from basic research, collections of genetic resources and industry. As users of genetic resources, these stakeholder groups were also assumed to have a potential interest in DSI.

In September 2017, the invitation to participate in the survey and the online link were distributed to various industry associations, research associations, research foundations etc. The contact details for these various organizations were provided by the German Federal Agency for Nature Conservation. The online link to the survey remained active for four weeks and was deactivated again in the middle of October.

It is not known exactly how many people received the invitation to participate or how many people accessed the survey link but ultimately decided not to complete the survey. As the survey was anonymous, it is also not known who completed the survey. Although the invitation to participate was directed specifically to German stakeholders, i.e. only distributed to organizations within Germany, the possibility that the link was forwarded to someone outside of Germany cannot be completely ruled out.

The online survey had 78 questions consisting of closed and open questions as well as statements where respondents had to indicate their agreement with a statement on a 5-point Likert scale, e.g. from very important to not important at all.

In the survey, the term “digital sequence information/data” was used and not the acronym DSI. No distinction was made between “information” and “data” for the purpose of the survey questions but both terms were included to avoid any possible confusion, e.g. which might arise by not using the term “data”, which is used commonly by the scientific community.

4. Results

4.1 The sample

4.1.1 The sample size

A total of 371 people completed the online survey and this level of response far exceeded initial expectations.

Those respondents who already work with DSI were regarded as relevant stakeholders and therefore only their views were included in the final analysis. This means that respondents who indicated that they do not currently work with DSI were removed from the dataset. Several respondents also indicated that they work exclusively with DSI from human genetic resources. As these do not fall within the scope of the Nagoya Protocol, their views were considered irrelevant and their responses were also removed from the data set. The responses of those respondents who indicated that they work with DSI from both human genetic resources and other genetic resources were retained. The final sample included 340 respondents. It is not known what percentage these respondents represent of the total number of people in Germany who use DSI.

Not all questions were answered by each respondent. Where no answers were provided, these responses were simply categorized as “not specified”. Answers to open questions were coded and analysed. Despite the length of the survey, up to two thirds of the respondents provided responses to the open questions. In a few cases, the answers provided to open questions were not clear and could not be analysed, or alternatively the answer provided was not relevant to the question. Where this occurred, those individual responses were simply excluded from the analysis of the relevant question.

4.1.2 The survey participants

Most of the respondents in the final sample completed the survey from their individual perspective. A few completed the survey on behalf of an association and/or its members and a further 26 respondents completed it on behalf of their employer/institution. The majority of the respondents (n=235) were men and over half of all the respondents were aged from 36 to 55. As might be expected due to the nature of the survey, the respondents were highly educated, with 45% of participants having completed a doctoral degree and a further 41% having completed a habilitation thesis.

Most of the respondents (n=293) indicated that they are employed in the public sector, with only a small number indicating that they are employed in either the private sector (n=14) or by public/private partnerships (n=18).

The respondents indicated that they are employed by (it was possible for respondents to allocate their employer or institution to more than one category):

- institutes or organizations focused on non-commercial research (n=301);

- institutes or organizations focused on teaching (n=86);
- collections of genetic resources (n=67);
- non-profit organizations (n=18);
- government (n=14);
- industry associations (n=5);
- institutes or organizations focused on commercial research (n=10);
- industry (production) (n=7); and
- organizations focused on biodiversity conservation (n=1).

Given that most respondents are employed in the public sector, it is not surprising that most respondents are employed by institutes or organizations that are engaged in non-commercial research. Around half the respondents (n= 174) indicated that their employer is exclusively involved in non-commercial research. There was, however, considerable overlap between institutions focused on non-commercial research and institutions with a teaching focus as well as between collections and institutions focused on non-commercial research. Overall, participation in the survey by people working on commercial research and in industry was very low.

In some cases, it was quite straightforward to determine whose institute/organization is commercial in nature, e.g. when respondents indicated that they are employed in the private sector and engaged only in commercial research and/or industry production. However, the boundary between non-commercial and commercial interests was not always clear. Several respondents indicated that their employer was involved in both non-commercial and commercial research and this included respondents from both the private and public sectors.

Unsurprisingly, many of the respondents currently occupy research related positions, including professors (n=117), researchers (n=77), principle investigator/lab head (n=50) and postdoctoral researchers (n=42). Other positions included lecturers, technicians, management and a science policy coordinator. Several students also participated in the survey.

Funding for the respondents' work and research comes from a range of different sources. As most respondents are employed in the public sector and are involved in non-commercial research, it is unsurprising that most are funded by federal (n=282), state/regional (n=137) or local governments (n=16). Some respondents (n= 92) receive funding from the European Union. Some of the respondents are funded by foundations (n=33), including private foundations and research foundations such as the German Research Foundation (Deutsche Forschungsgemeinschaft). Some of the respondents receive funding from corporations or industry (n=29) and some of this private funding flows to respondents who indicated that they are in the public sector and involved in non-commercial research. One person indicated that they receive funding from an animal breeding organization. Several people receive

funding from non-profit organisations and a couple of respondents indicated that they are self-funded.

The respondents who use DSI are involved in a range of different disciplines, most of which would be characterised as being part of the life sciences. These included:

- agronomy
- algal diversity
- animal behaviour
- animal breeding
- animal health
- biochemistry
- biocontrol
- bioinformatics
- biology
- biomimetics
- biotechnology
- botany
- chemistry
- computer science
- developmental biology
- ecology
- engineering
- entomology
- evolutionary biology
- food
- genetic engineering
- genetics
- genomics
- health sciences
- identification of invasive and pest organisms
- immunology
- infection biology
- information science/technology
- integrative physiology
- medicine
- metabolomics
- metagenomics
- microbiology
- molecular biology
- molecular systematics
- neurophysiology
- neuroscience
- pharmacy
- phenomics
- phylogenetics
- phylogenomics
- plant breeding
- plant pathology
- plant physiology
- plant science
- population genetics
- proteomics
- soil science
- synthetic biology
- taxonomy
- transcriptomics
- zoology

Approximately 80% of the respondents indicated that they are involved in more than one discipline. The most common fields of research were more traditional fields of biology such as evolutionary biology (n=147), taxonomy (n=110) and ecology (n=89). Some respondents also indicated that they are working in the newer “omics” disciplines such as genomics (n=99), transcriptomics (n=46) and metagenomics (n=27). A complete list of the research disciplines and the number of respondents in each discipline is provided in Annex A.

Only a small number of respondents (n=16) specified in which commercial/ industry sector they are employed. These industry sectors included:

- animal breeding;
- biotechnology;
- food and feed;
- pharmaceuticals;
- plant breeding; and
- other sectors (DNA services, medical devices).

Several respondents indicated that they belong to more than one sector, e.g. both biotechnology and plant breeding. Two respondents indicated that they are involved in industry sectors which were not specified in the questionnaire, namely DNA services and medical devices. There were no respondents from the cosmetics or biocontrol and biostimulants sectors in the final sample.

A further five respondents indicated that they are somehow related to certain industry sectors, including biotechnology, pharmaceuticals and plant breeding, although they are employed in non-commercial research. It is not possible to discern exactly what the relationship between these respondents and the various industry sectors might be. It could be that the findings of their non-commercial research inform industry in some way or could be applied in the future or there may be collaboration with industry actors.

4.2 The use of DSI

All 340 respondents included in the final sample currently use DSI and almost all (n=338) indicated that they expect to use it in the future. Most of the respondents use DSI generated through their own work as well as DSI from other sources, i.e. information/data generated by people outside their own working groups or collaborations. Only a few respondents (n=6) indicated that they work exclusively with their own DSI and some people (n=32) indicated that they are working exclusively with DSI from other sources.

The DSI with which the respondents work comes from a range of different organisms as shown in Table 1, with the most common sources being animals, bacteria and plants. Just under half of the respondents (n=164) indicated that they work with DSI from only one type of organism.

Table 1 Types of organisms from which DSI is obtained.

Types of organisms	No. of respondents
Animal	218
Bacteria	156

Types of organisms	No. of respondents
Plant	113
Fungi	85
Archaea	63
Virus	65
Algae	57
Protists	3
Protozoa	3
Metagenomes	2
Amoeba	1
Not specified	2

4.2.1 Use of synthetic genetic resources

It is possible for some researchers to generate synthetic genetic resources using sequences and a DNA synthesizer (Scott et al., 2015). The survey attempted to determine how many respondents are, at least theoretically, in the position to work with synthetic genetic resources. Comments made by a couple of respondents indicated that the related questions in the survey instrument were problematic for them because no definition of "synthetic genetic resources" was provided. As such, these results should be treated with some caution as the question may not have been understood by the respondents in the way it was intended by the author.

Most respondents (n=209) indicated that it would not be possible for them to generate synthetic genetic resources in their own laboratories but around half the respondents (n=178) indicated that it would be possible to have another laboratory to do it on their behalf. Although the possibility exists for synthetic genetic resources to be created, over three quarters of the respondents indicated that they would not, even theoretically, be able to work exclusively with synthetic genetic resources in future. A couple of respondents commented that the generation of synthetic genetic resources is not always feasible and can be more time consuming and expensive than using genetic resources.

Most of the 69 respondents who indicated that it might be theoretically possible for them to work only with synthetic genetic resources in future are still generating sequences through their current work, i.e. working with genetic resources. One respondent indicated that he already works with synthetic genetic material (in combination with other types of genetic resources) and several other respondents commented that they use DSI for gene synthesis.

4.2.2 Source of the original genetic resources

The respondents in this study do not always know where the genetic resources from which the DSI was obtained come from. As shown in Figure 1, less than one third of respondents indicated that they always know the origin of the genetic resources. Some respondents even indicated that they never know (n=15) or rarely know this information (n=28).

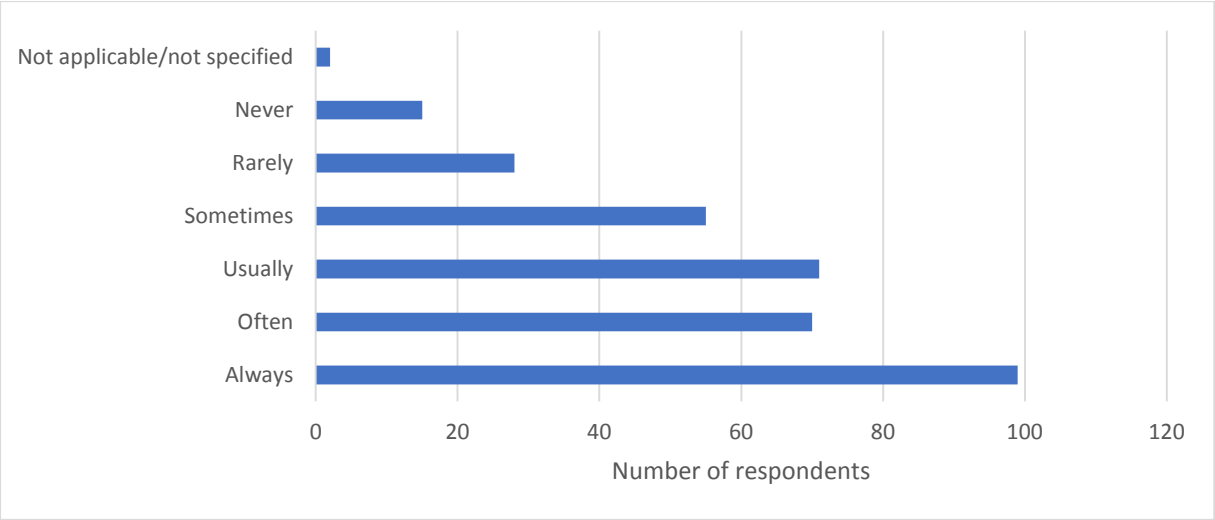

Figure 1 Frequency with which respondents know where the original genetic resources underlying the sequence come from.

Reasons provided by the respondents as to why the origin of the genetic resources is not always known included:

- records about the origin of the genetic resources simply do not exist, the samples of sequenced genetic material are too old or the information has been lost;
- there are no metadata accompanying sequences about the origin of the respective genetic resources;
- the metadata accompanying sequences are inexact or unreliable;
- the respondents have no access to the relevant metadata;
- the origin of the genetic resources is recorded in metadata in the database or is provided in the underlying publications but the respondents do not check this information;
- the quantity of data being used makes it impossible to keep track of the source of the genetic resources for every single sequence, i.e. it is simply not feasible for large datasets; and
- the respondents are not interested in the origin of the sequence, i.e. it is irrelevant to their work.

Comments made by respondents indicated that for some species, such as endemic species or species with limited distribution, it may be possible to identify the origin of the genetic resources even without having the relevant metadata, whereas for common species or widely distributed species, this would not be possible. Several respondents also suggested

that microbes and pests are often globally or at least very widely distributed and for this reason, it is difficult to reduce their origin to a specific region or country.

Some respondents commented that the origin of the genetic resources is entirely irrelevant for their research. Other respondents indicated that the origin is very important information for their work and a couple of respondents commented that it is their institutional policy that sequences cannot be used if the origin of the underlying genetic resources is not known.

Figure 2 shows that the genetic resources from which the DSI was generated come from all over the world (where this information is known).

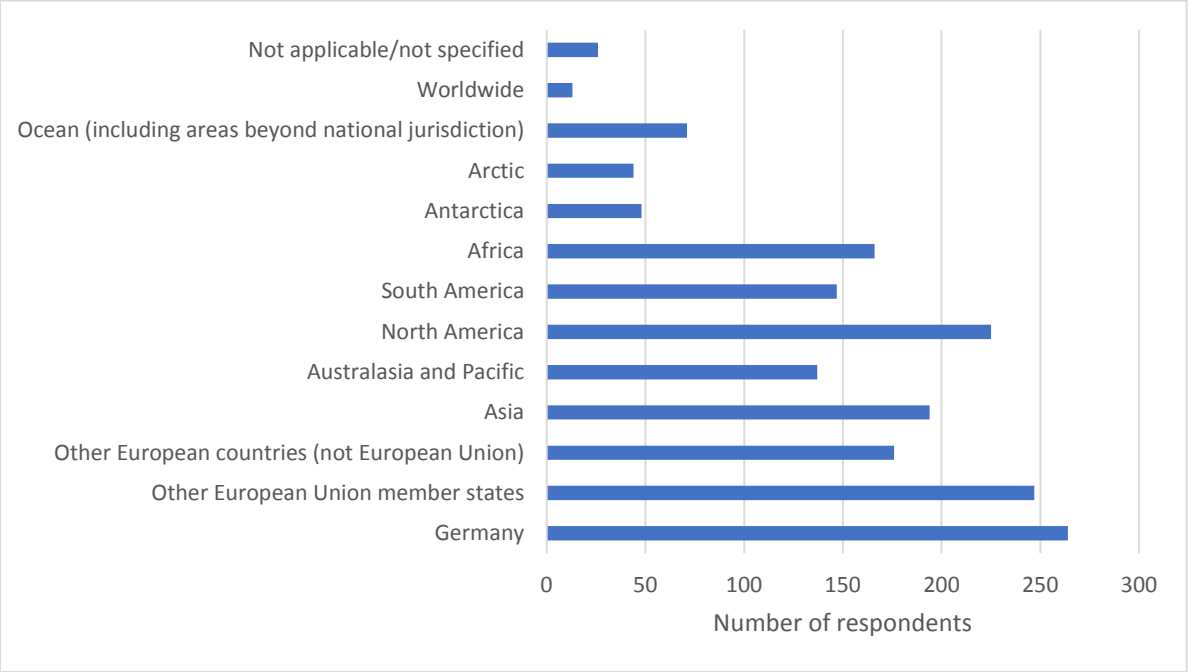

Figure 2 Origin of genetic resources underlying the digital sequences (in cases where known).

Many respondents indicated that they work with DSI that comes from genetic resources from Germany. A small number of people (n= 6) even indicated that they work exclusively with DSI from organisms that come from Germany. This needs to be treated with some caution because it is not clear whether these organisms come from Germany or were simply sourced from collections of genetic resources located in Germany.

The respondents work with DSI that is derived from organisms from various regions of the world. Given the global distribution of these genetic resources, it seems likely that at least some of this DSI is obtained from organisms that originally come from countries that are party to the Nagoya Protocol.

4.3 DSI generated through the respondents' own work or research

4.3.1 Volume of DSI generated through the respondents' own work or research

Most respondents (n=298) indicated that DSI is generated through their own work or research, including data generated directly by them, by another laboratory on their behalf or by someone within their research group/collaboration.

Respondents indicated that they generate different numbers of sequences through their work. Over one third of the respondents indicated that they generate less than one hundred sequences per year and around 20% indicated that they generate hundreds of sequences each year. Other respondents indicated that they generate very large numbers of sequences annually, i.e. thousands (n=58), millions (n=55) and billions (n=17). When asking for the number of sequences generated through the respondents' work, the survey instrument did not specify any sequencing units (e.g., basepairs, megabases, sequence length, gene number, etc.) and for this reason, these numbers are not directly comparable. It is possible that some respondents are referring to a whole sequence, e.g. a gene sequence, whereas other respondents may possibly be referring to the number of base pairs generated, which can range anywhere from hundreds (e.g. gene), tens of thousands (e.g. virus genomes), millions (e.g. bacteria genomes) to many millions of base pairs (e.g. plant genomes or environmental samples containing a mix of nucleotides from thousands of different microorganisms).

As shown in Figure 3, almost three quarters of the respondents indicated that the amount of DSI generated through their work has been increasing in some way since 2012, which was chosen arbitrarily as this was the year the Nagoya Protocol came into force. Under 20% of the respondents indicated that the amount of DSI being generated has stayed constant and around 6% indicated that the number of sequences generated through their work decreased during that period.

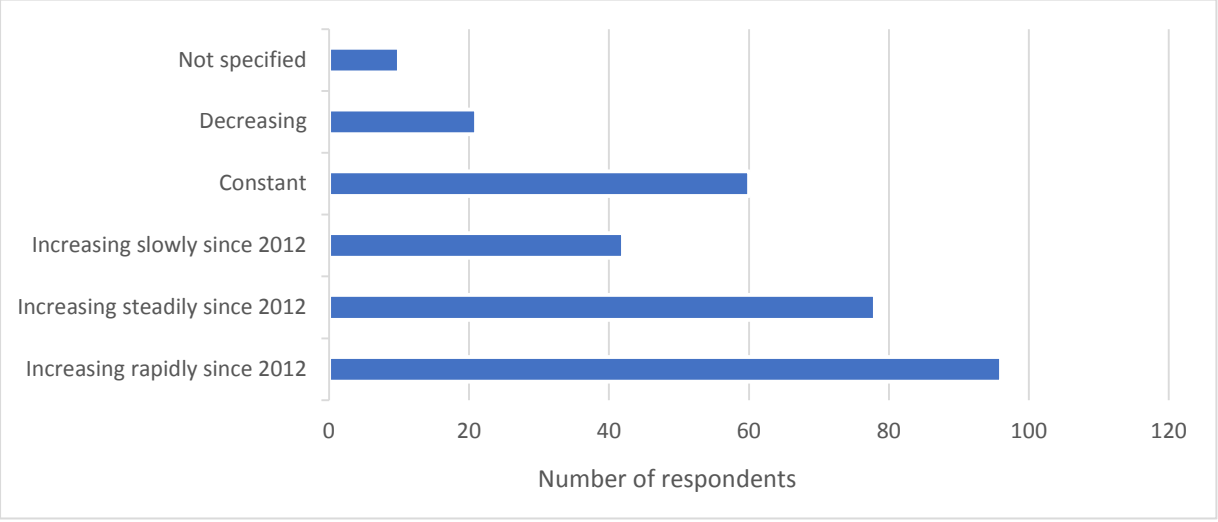

Figure 3 Trend in amount of DSI generated through the respondents' own work or research.

4.3.2 How the respondents use the DSI generated through their work or research

The respondents were asked to give a short description of how they use the DSI generated through their own work. It was not expected that the complexity of their work could be reflected in detail and for this reason, it was intended only to provide a general overview. Most respondents (n=280) provided an answer to this question and they reported on a range of different uses. The responses can be divided into general and more specific answers.

General uses included answering research questions, testing hypotheses, using it for comparative purposes and for experimental design as well as for publication purposes. One respondent indicated that DSI is used for reporting to governmental administrative agencies. Some respondents use DSI for educational purposes.

Most respondents provided more specific responses about how they use DSI. It is employed in a very diverse number of disciplines and for very different purposes. No attempt has been made to determine whether these uses could be regarded as "utilization" in the sense of the Nagoya Protocol.

DSI generated through the respondents' work is used for:

- taxonomic purposes, e.g. to identify taxa, to compare sequences between related taxa and to analyse similarities. DSI is also used to identify species and to describe new species. Some respondents generate DNA barcodes;
- phylogenetics, including for the generation of phylogenetic trees and phylogenetic reconstructions;
- the study of evolution, for example, for investigating evolutionary relationships, the rate of evolution, evolutionary processes, drivers of evolution, adaptation and gene evolution. Some respondents investigate the processes shaping biodiversity and sequence variability within species;
- ecological studies, e.g. to investigate populations, population structure, population genetic analysis, the relationships between populations and the distribution of populations in space and time; and
- research on plants, including basic research, sequencing of plant genomes, investigation of plant characteristics and improvement of plant growth as well as investigating plant evolution, e.g. the domestication of cereal crops.

DSI is used by respondents in their work with microorganisms, including:

- to identify microorganisms, e.g. pest species, strain identity, new microbial species etc.;
- to compare sequences to identify mutants;
- for investigating microbial communities and their diversity;
- for identification required for deposit in culture collections;
- for investigating pathogens;

- to investigate regulatory mechanisms;
- for microbial cell biology; and
- in metagenomics to look for useful enzymes.

Other examples of how the DSI generated by the respondents is used include:

- identification of new genes and gene clusters;
- identification of gene loci;
- identifying the relationship between genes and mutations (sequence comparisons);
- identification of sequence polymorphisms;
- investigating gene expression and identifying gene function;
- understanding the link between genotypes and phenotypes;
- primer design;
- cloning;
- targeted mutation of sequences or creation of new functional sequences, e.g. for applied use in synthetic biology;
- genetic engineering;
- screening and verification of genetically modified organisms (GMO);
- understanding how physical regulation is reflected by molecular biology;
- investigating the physiology of cells and organisms;
- analysing transcription under different conditions;
- investigating epigenetics;
- investigating horizontal gene transfer;
- using DSI as template for protein production;
- investigating the production and function of proteins;
- work with recombinant proteins, e.g. characterisation and use for the synthesis of chemicals;
- analysis of basic metabolic function, metabolic and biosynthetic pathways, searching for new biocatalysts as well as metabolic engineering;
- characterizing enzymes;
- mining for enzymes and other compounds;
- creating novel enzyme catalysts;
- paternity assignment;
- molecular breeding support and optimization of breeding;
- determining the sex of organisms, e.g. birds;
- investigating disease/epidemiology;
- the development of diagnostics;
- quality control/biosafety assessment; and
- deriving mathematical models or training machine learning models.

4.3.3 Sharing and publishing DSI

It was assumed for this study that there is a difference between sharing data and the publication of data and the questions in the survey instrument reflected this assumption. However, it was apparent from the responses provided that not all respondents recognized this difference, meaning that the survey instrument could have been improved by stating the difference explicitly. Due to the confusion, those people who indicated that they had only "shared" information through public databanks, i.e. published it, were not included in the analysis of this question.

Some respondents (n=34) indicated that they have never shared their DSI but most respondents (n=254) indicated that they have. Of those respondents who have shared their DSI, approximately half of them have shared 50% or more of their DSI, as shown in Figure 4.

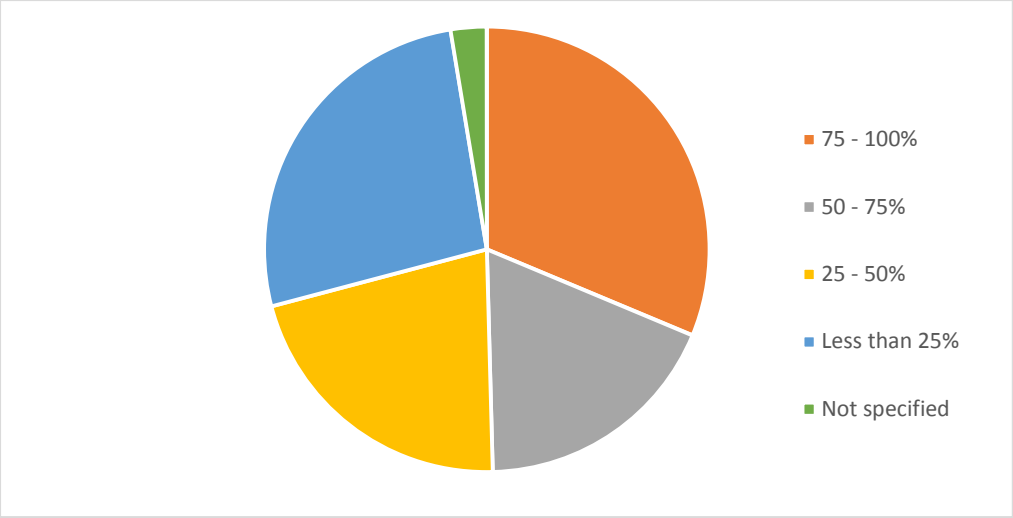

Figure 4 Percentage of DSI shared by the respondents.

DSI is shared in various ways but most commonly (n=168) by way of direct contact/email. Other specific sharing mechanisms included:

- networks;
- personal or institutional websites;
- file hosting services;
- file transfer protocols or other webtools operated by the individual institute;
- clouds; and
- hard disc.

Some respondents (n=23) indicated that they have never published any of their DSI. Most (n=272), however, have published at least some of their DSI. Just over half the respondents indicated that they have published 50% or more of their DSI. Around 13% of the respondents indicated that they have published less than 25% of their data. Although additional comments made by a few respondents indicated that they publish everything, it

seems that in general, not all DSI generated through the respondents' work becomes available to the public.

As shown in Figure 5, the respondents publish their DSI in different ways but the majority use publicly accessibly databases for this purpose.

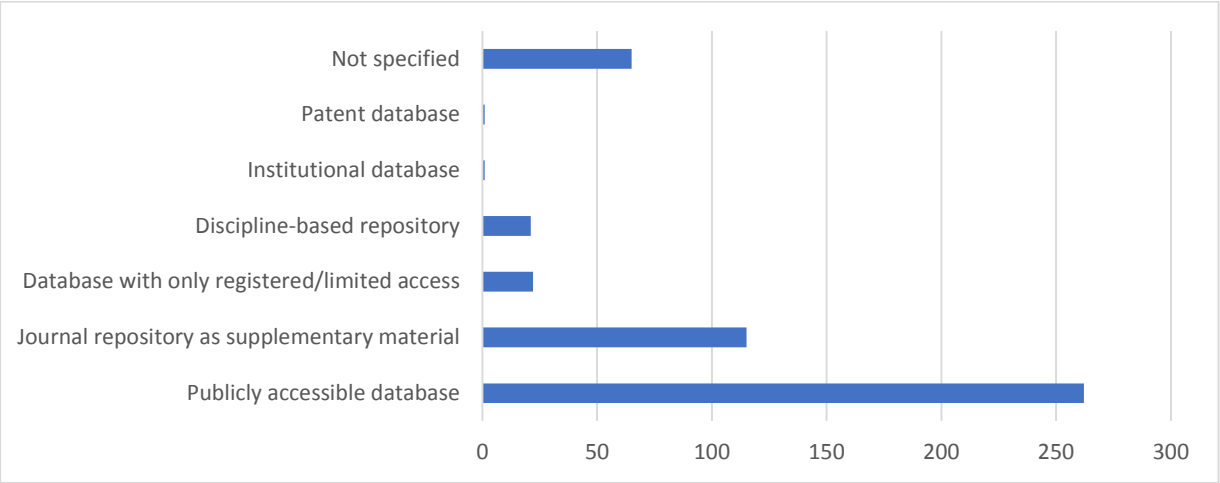

Figure 5 Places where respondents publish their DSI.

A number of respondents reported that they upload their DSI to more than one database. The respondents use various databases but most people upload their DSI to databases held by the NCBI and EMBL-EBI, with NCBI's GenBank being by far the most commonly used (n= 215). A full list of the various databases, platforms etc. where respondents make their data available is provided in Annex B.

Once DSI has been generated by the respondents, it is not always immediately available to others. Access may be restricted to certain actors including the individual researcher (n=17), the research team (n=228), collaborators and cooperation partners (n=176), the institute/employer (n=75) and the sequencing laboratory (n=62). Few of the respondents publish their DSI immediately after sequencing. Most of the respondents (n=187) publish or share their DSI after their results have been published or at the time of publication. Respondents provided various reasons for why DSI is published at different times. Some respondents indicated that this can stem from intellectual property rights or the time needed for processing the data and checking data quality. Institutional policy also plays a role for some respondents. One respondent commented that access to DSI generated through their work is restricted to the research consortium for a fixed period of several years before it can be published and another respondent indicated that when research is conducted in collaboration with corporations, i.e. the work is privately funded, the data is not published.

Most respondents indicated that it was very important to be able to both share their DSI with others and to be able to publish it, as shown in Figure 6.

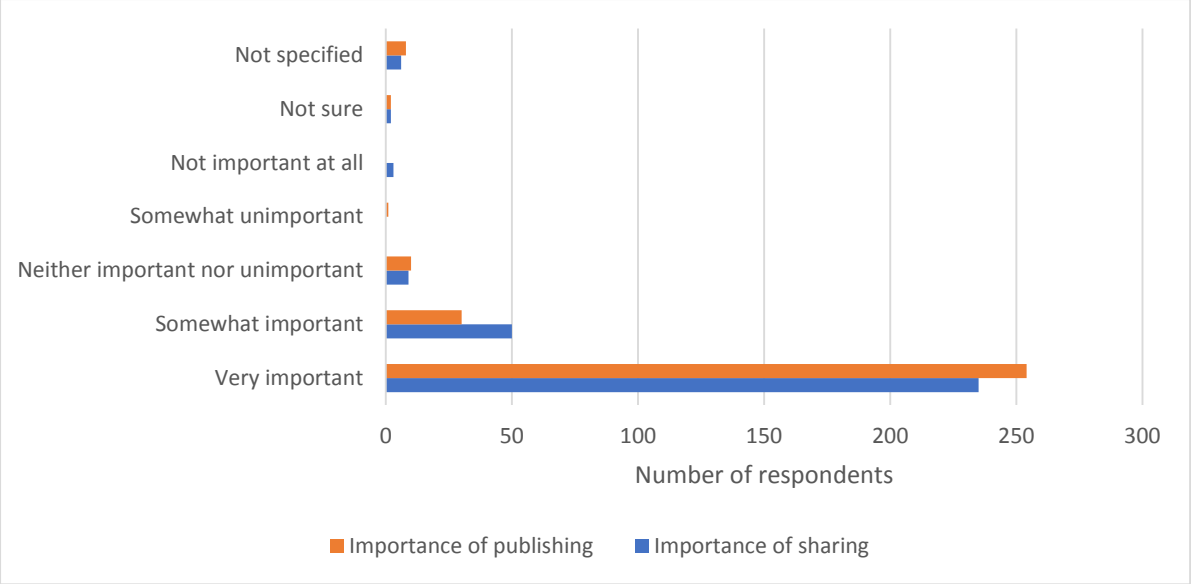

Figure 6 Importance of being able to share and publish DSI.

As shown in Figure 7**Fehler! Verweisquelle konnte nicht gefunden werden.**, some of the top motivators for sharing or publishing data were that it is standard practice of the scientific community, publishers require it, it contributes to the acceleration of research as well as their personal commitment to open data.

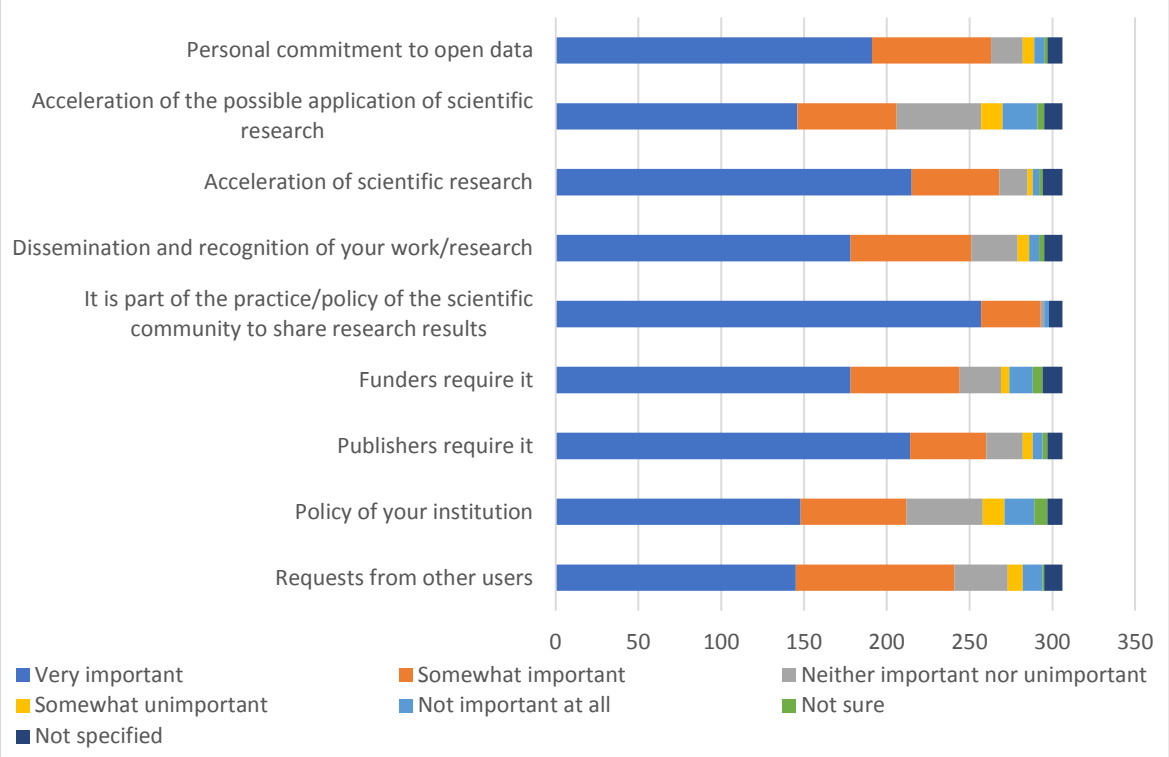

Figure 7 Importance of selected motivators for sharing or publishing DSI.

Some of these motivators for sharing or publishing data might be regarded as internal motivators whereas others are more external motivators, which are beyond the control of the individual researcher. One respondent commented, however, that “motivation” is not

an appropriate term when referring to the publication of data because science is evidence based, evidence has to be documented and made available to others and thus publication is a “rule” and not a question of motivation.

Other reasons for sharing and publishing DSI which emerged the respondents’ comments included:

- the data are generated using public funds and therefore should be published;
- the opportunity to collaborate and network with others, including at the international level;
- friendship and the enjoyment of working with others;
- the perceived benefit to the wider community from data sharing;
- the ability to check data and improve the quality of research;
- sharing data allows research to be reproduced;
- data sharing sets a good example for young scientists;
- ethical reasons;
- reciprocity, meaning that by sharing their own data, respondents want to encourage others to their share in the hope that they will somehow have an advantage in the future by having access to other people’s data;
- efficiency, i.e. it makes research faster and saves money by avoiding the replication of work;
- it reduces the need for animal experiments;
- taxonomy is particularly dependent upon collective efforts; and
- rapid identification of organisms is possible, including identification of potentially harmful pathogens.

4.3.4 Concerns about potential liability arising from sharing or publishing DSI

When asked about potential concerns regarding liability arising from sharing or publishing their DSI, most respondents (n=228) indicated that potential liability was not a concern for them. Several respondents indicated that they had never thought about the possibility of liability. Respondents were asked to provide the reasons for their lack of concern about liability. These included:

- their work is non-commercial, meaning the researcher has no commercial interest in the data or alternatively the data have no potential commercial value, e.g. barcode data or phylogenetic data;
- there is no apparent way for their data to be misused, e.g. sequences are not potentially harmful;
- third parties bear the responsibility for how they use published data;
- respondents have obtained all relevant permits prior to sharing or publishing data;
- all intellectual property rights are clarified prior to starting projects;
- the material from which the sequences were obtained is not patented;

- they do not try to exert any patent rights themselves;
- the DSI they work with is not hard to access, i.e. other researchers could easily get access to the same data;
- they work in a transparent way;
- they are not working with human sequences;
- publishing is normal practice in science;
- their data is stored in publicly accessible databases;
- their data should be open and freely accessible to anyone;
- publishing data is required for publication;
- it is their institutional policy to publish;
- results generated with public funds have to be published; and
- the belief that no one owns the sequences.

Around 20% of the respondents indicated that they do have concerns about liability and at least some of this concern seems to stem from the Nagoya Protocol. Several respondents (n=11) indicated that the Nagoya Protocol has created uncertainty about what they are and are not permitted to do. Several respondents stated that they actively aim to comply with relevant Nagoya regulations but they still have concerns about compliance.

Other reasons for having some concern about liability included:

- administrative regulations and law generally;
- intellectual property rights;
- the possibility that third parties might do the wrong thing. A couple of respondents indicated that once published, they cannot control what is done with their data and therefore it is not possible to rule out misuse, e.g. commercialization or republishing without consent; and
- sequences are potentially dangerous, e.g. in the case of pathogens.

4.3.5 Potential for tracking and tracing access to and use of published DSI

Most respondents indicated that once their DSI has been published, e.g. in publicly accessible databases and repositories, it is not possible to know if it is accessed or how it is used by others, as shown in Figure 8.

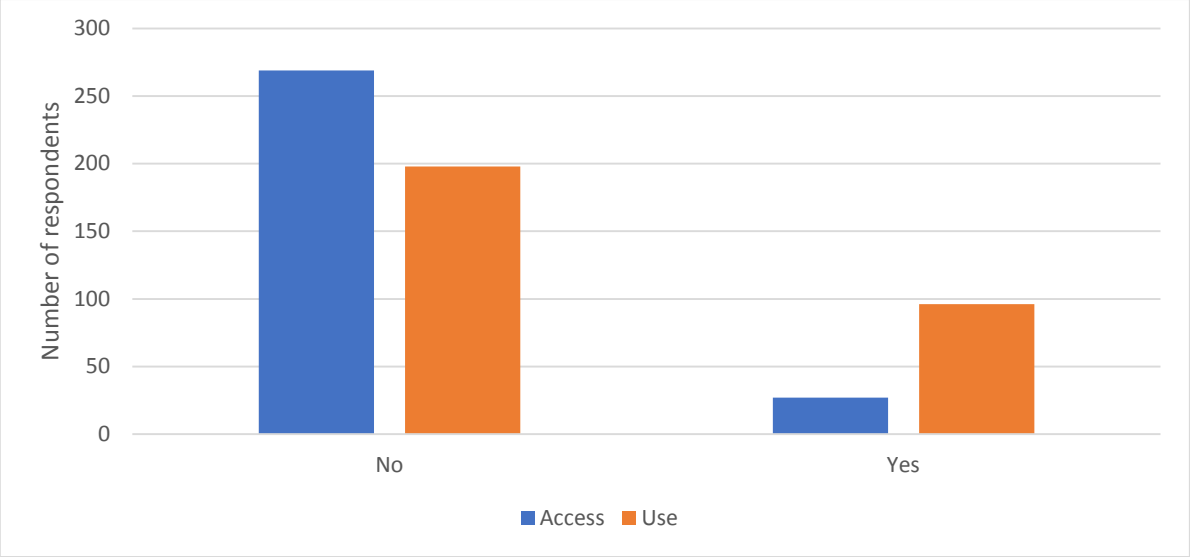

Figure 8 Knowledge about access to and use of the respondents' published DSI.

Most participants commented that tracking and tracing access to published DSI is not possible. A number of respondents noted in their comments that once DSI is published in public databanks, repositories, archives or in journal repositories, it is accessible to anyone and they no longer have any control over it.

There seemed to be mixed opinions about whether public databases and repositories etc. can trace access to data. Most respondents indicated that these databases do not trace or log downloads, meaning that the downloads are anonymous. A couple of respondents also indicated that databases are often downloaded and searched locally, meaning that the databases themselves have no information about the individual searches, the search results or what is done with these results. Other respondents suggested that repositories may have algorithms that enable tracking and tracing but even if this is possible, this information is not disclosed to the public or to data submitters. Some respondents also suggested that tracing, even if possible, would not be feasible due to the number of downloads or the volume of data.

Several respondents indicated that it may be possible to know about access/use of published DSI through direct personal contact and through collaborations. Otherwise, respondents indicated that they can infer access to DSI and learn about subsequent use if the people using the data also publish their work. This is, however, dependent upon the subsequent users providing correct citations and/or using the sequence accession numbers. Some respondents indicated that this can be problematic as citation is not always done properly. Even if they are cited, some respondents also suggested that they may not become aware of this unless they conduct a manual search of citations and that searches of this nature would be resource (time) intensive and of little interest. A couple of respondents also pointed out that not all subsequent uses of DSI are published, especially those in the private sector, and thus published work only provides an incomplete picture of who has accessed DSI and how it has been used.

Although a couple of respondents indicated that it would be interesting to know about subsequent work with their published DSI, more respondents expressed no desire to know about access or further use of their published DSI.

4.3.6 Reuse of published DSI

Despite the apparent difficulty with knowing who accesses and uses their published data, around three quarters of the respondents reported that they know that their published DSI had been reused in some way. While most indicated that their DSI had been incorporated into larger data sets (n=181), reanalysed (n=125) and used to draw new conclusions (n=135), some reported that it had been used for other purposes such as developing or checking new methodologies, cloning, gene annotations, testing different hypotheses, and educational purposes. A small number of respondents also indicated that their published data has been used for the development of products (n=14).

4.3.7 Importance of generating DSI

For most of the respondents, the ability to generate DSI through their own work is important. Around 84% said that it was very important and a further 11% indicated that it was somewhat important, as can be seen in Figure 9.

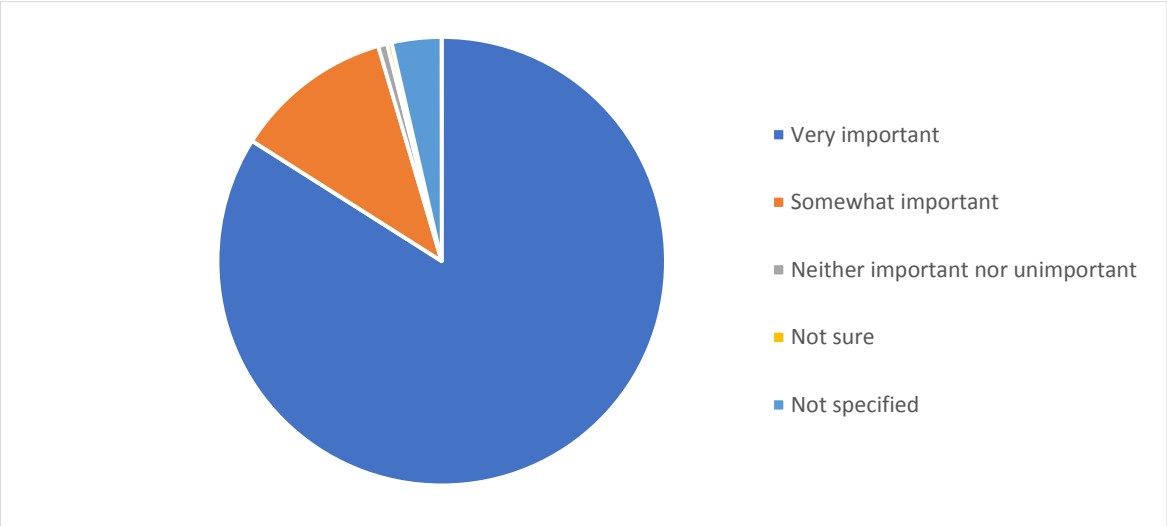

Figure 9 Importance of DSI to the respondents' own work.

For many respondents (n= 284), it would be impossible to do their research or work without generating DSI. Respondents provided a broad range of reasons for this, including:

- it is their focus or a core aspect of all or part of their own work/research;
- it is the focus or core aspect of the business that they work for;
- it is essential for the types of analyses or experiments being conducted;
- it provides the only way to get the desired information;
- it has become a standard tool, meaning it would not be feasible to work without it based on the current state of knowledge;

- it saves both time and resources. One respondent even suggested that without DSI, research in Germany would not be competitive with other countries, e.g. the USA;
- use of traditional methods would be a major backward step;
- they are on the only researchers/group generating sequences for a particular species;
- generation of new DSI is an important part of discovering new species;
- there would be potential disadvantages associated with relying exclusively on other people's data, e.g. because of data quality issues, which makes it necessary to generate their own DSI;
- legal reasons, e.g. the identification of strains; and
- generating DSI through their work is essential for publication purposes.

Examples of fields where respondents indicated it would be impossible for them to work without generating DSI were phylogenetics, evolutionary biology, taxonomy, molecular biology, genetics, genomics, bioinformatics, molecular systematics, biochemistry, biotechnology and genetic engineering, and work with microorganisms.

Only around 5% of the respondents indicated that it would be possible to continue their work without generating DSI. According their comments, this would be possible because:

- sequencing is only one aspect of their work;
- DSI adds further insight into their work but it is not the only form of data that can be used, e.g. it would be possible for them to work with morphological data;
- they could work with more traditional methods; and
- in vivo experiments would be unaffected.

4.4 Use of DSI from other sources

4.4.1 Volume of DSI from other sources that is used

Most respondents (n=328) indicated that they use DSI from other sources, i.e. DSI generated by people outside their own working groups or collaborations. Some respondents (n=32) indicated that they work exclusively with DSI from other sources.

The volume of DSI from other sources which is used by the respondents each year ranges from less than one hundred sequences up to billions of sequences annually. Approximately half the respondents indicated that they use hundreds of sequences or less than one hundred sequences from other sources. A quarter of the respondents indicated that they use thousands of sequences and around 20% of the respondents indicated that they are working with very large volumes of data, i.e. millions (n=51) or billions (n=17) of sequences. Again, it should be kept in mind that no sequencing units were provided in the survey instrument when asking for the number of sequences from other sources that are used

annually and it is thus not clear whether respondents are referring to complete sequences or the number of base pairs.

Around 70% of respondents reported some sort of increase in the amount of DSI from other sources which they are using annually, as shown in Figure 10. Comments made by one respondent indicated that, at least in the case of their organization, the increase in data use is exponential rather than linear.

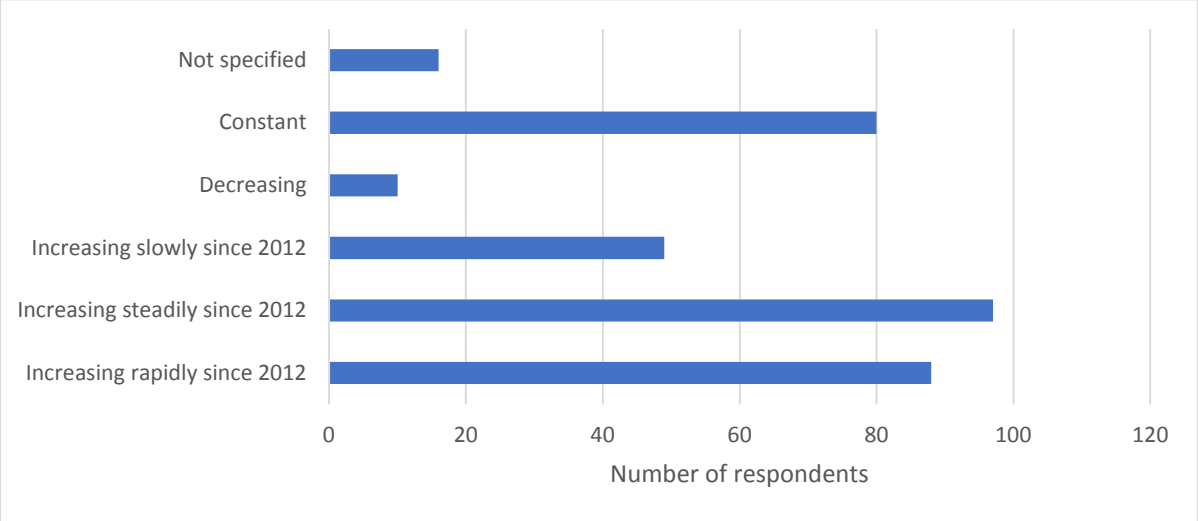

Figure 10 Trend in usage of DSI from other sources.

4.4.2 Sources of DSI

Respondents obtain their DSI from a range of different sources, as shown in Figure 11. The most common source of DSI is publicly accessible databases but many people also obtain DSI by way of email/personal contact and through journal repositories. It is not surprising that respondents obtain their DSI through these avenues as many of them share and publish their own data in the same way.

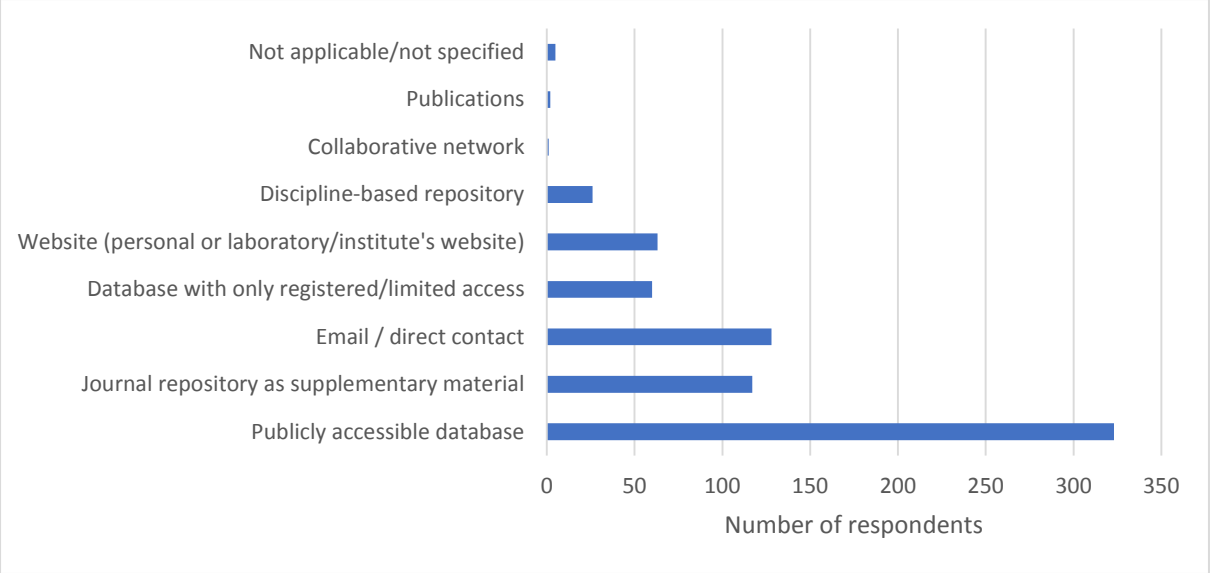

Figure 11 Places where respondents get DSI that has been generated by others.

The respondents use various databases to obtain DSI but again NCBI's GenBank was by far the most commonly used (n=265). A full list of the databases/platforms etc. where respondents get their DSI is provided in Annex C.

4.4.3 How DSI from other sources is used

The respondents were asked to give a short overview of how they use DSI obtained from other sources. Most respondents (n=309) provided a short description of how they use it. Again, the purpose of this question was not to capture the complexity of the work done by the respondents in detail but to gain a rough overview. Unsurprisingly, respondents use data from other sources in a similar way to how they use their own data.

The responses can be divided into general and more specific answers. Generally speaking, the respondents use DSI obtained from other sources:

- to further their research, generate hypotheses and test ideas;
- to inform experimental design;
- to improve laboratory methods;
- to expand their own dataset, i.e. they add it to the data generated through their own work or enhance their own dataset with it. Some respondents commented that this allows them to address new questions, increase the robustness of their findings etc.;
- for comparative purposes, e.g. as a reference to which they can compare their own data, to correct false interpretations, for verification etc.;
- for re-analysis;
- for meta-analyses; and
- for educational purposes.

Examples of specific uses of DSI from other sources included:

- phylogenetic analyses and generating or reconstructing phylogenetic trees. Respondents commented that using data from other sources is practical due to the large volume of data needed, and that it allows results to be checked and mistakes to be corrected etc;
- taxonomic work, such as investigating the relationships within and between species or taxa, to validate previous research and to generate new taxonomic knowledge;
- tracing evolutionary patterns;
- genetic analysis of populations and biogeographic analyses;
- comparative genomics;
- homology searches;
- alignments, e.g. using the Basic Local Alignment Search Tool (BLAST) to find regions of local similarity between sequences, to check one's own sequences, to detect new species and functions as well as for contamination control;
- for primer design and generation;
- to identify metagenomic sequences;
- to identify and compare genes, investigate gene structure, to obtain information on gene loci, to build gene models and investigate putative gene function and gene evolution;
- for analysis of gene expression;
- for constructing novel gene expression sequences;
- for annotation purposes;
- for single nucleotide polymorphism detection and analysis;
- to test environmental effects on epigenetic modification;
- to investigate mutation;
- to create new DNA constructs;
- for investigating metabolic pathways;
- to investigate the structure, function and interactions of proteins;
- for the identification and characterization of enzymes;
- for cloning purposes;
- for synthesis of genes and proteins;
- for investigating physiological changes;
- for genetic engineering;
- to create new diagnostic tools and probes; and
- for building mathematical models and bioinformatic workflows.

Some respondents (n=103) indicated that they have republished DSI after using it. Of those respondents who indicated that they are employed in industry and commercial research, several (n=5) indicated that they have used DSI from other sources for product development. These respondents obtain their DSI from a range of different sources, including different types of databases, journal repositories and through direct contact.

4.4.4 User agreements or conditions on the use of DSI from other sources

Over half the respondents (n=201) indicated that they are not aware of any conditions/user agreements associated with using DSI from other sources. Over a third of the respondents (n=125) reported that they are aware of various conditions, which are shown in Figure 12.

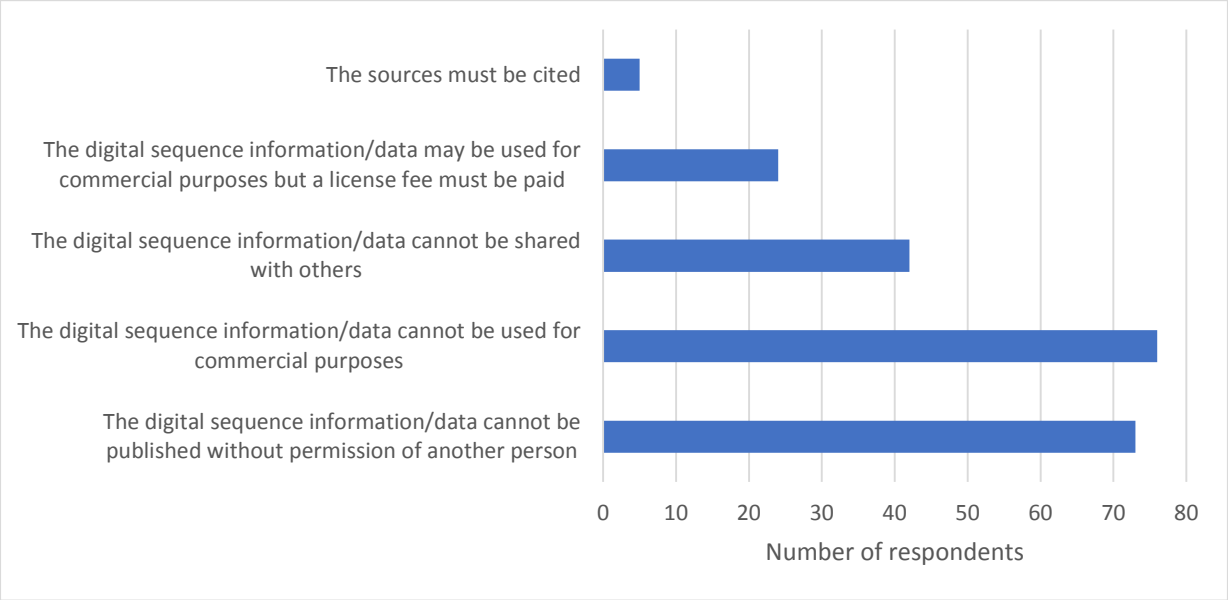

Figure 12 Conditions on using DSI from other sources.

Comments made by some respondents indicated that the presence of conditions will depend largely upon the specific database or repository, will not apply to all data and are most likely to apply to data from databases which require registration. One respondent indicated that it is his understanding that all data published in public databases such as GenBank are free to use without any form of restriction.

4.4.5 Importance of having access to DSI from other sources

Having access to DSI from other sources appears to be very important for the respondents. Most indicated that having no access to DSI from other sources would be a major impediment for their work/research (87%) and would restrict their work/research (85%). Only a small number of respondents indicated that not having access to other people's DSI would not be a problem for them.

The extent to which respondents rely on DSI from other sources varies. Some respondents cannot work exclusively with DSI but regard it as a very important tool that is used routinely in their work or research. Some suggested that without DSI from other sources, they would have to spend a large amount of time sequencing data themselves, which several respondents stated would simply not be feasible. The DSI available in databases has been collected over a long period of time by many people and respondents suggested that it would not be possible for individuals to generate this information again for the purpose of their own research. It was also suggested that it would be a considerable waste of resources for activities to be repeated if sequences have already been described.

There were respondents who already rely entirely upon DSI from other sources, including respondents involved in both non-commercial and commercial research. In some cases, the ability to work exclusively with other people's DSI is limited only by the availability of the desired data. A few respondents indicated that if all the DSI they needed were available in publicly accessible databases, it would be possible for them to work using only other people's DSI. A couple of respondents suggested that they are almost at the point where it is entirely possible to do their research using DSI from other sources and that only in exceptional cases they would need to sequence genetic resources themselves.

Several respondents indicated that although it would be possible for them to work exclusively with DSI from other sources, this would potentially restrict their research, e.g. to specific research questions. Examples provided by respondents of those types of analyses which would be possible using only DSI from other sources included *in silico* analyses, some comparative analyses and evolutionary interpretations, some aspects of taxonomic research and phylogenetic analyses as well as some analyses of protein/enzyme encoding sequences.

Most of the respondents (n= 261) indicated that they could not work exclusively with DSI from other sources. They provided various reasons for this, including:

- the researchers lack the necessary skill to work only with DSI;
- there is too little data in databases, specifically:
  - analysis would be too restricted based on the data that is currently available;
  - the sequences of interest are not available, e.g. there has been insufficient sampling or coverage of the organisms of interest, meaning that they have to generate their own data;
  - the researchers are the only ones working on that organism and therefore all data comes from their own working group;
  - information/data does not exist for new species;
- simply limiting research to reanalysis and meta-analysis of other people's data would be possible but would not progress their field of research;
- the quality of research is improved by having more data;
- the available data may be limited in some way, e.g. the data may not be reliable, correct or complete etc.;
- DSI is just one aspect of work and these data need to be combined with other types of data;
- their work necessitates use of DSI and a physical sample such as tissue or the whole organism, e.g. for wet lab experiments or to investigate function or the link between phenotypes and genotypes;
- DSI cannot answer specific research questions; and
- there are limitations to what can be tested *in silico*, e.g. responses to certain environmental conditions cannot be tested using DSI.

A number of respondents also expressed a desire to have original data and investigate something new.

4.4.6 Access to DSI from other sources

Given that lack of access to DSI would be an impediment or hindrance to the work of most of the respondents, it is not surprising that most of them agreed strongly with easy, unrestricted access to data that is free of charge and which may be used without restriction, as shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**. There was less support for the statement that "Access to DSI is low cost".

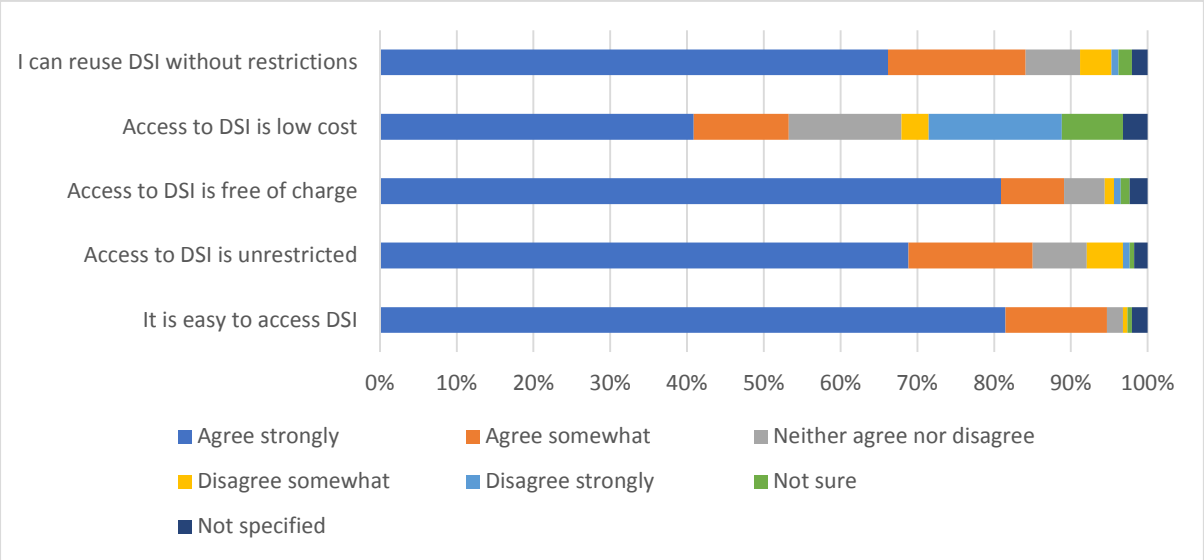

Figure 13 Level of agreement with statements about access to and reuse of DSI from other sources.

A number of respondents made comments specifically with respect to open access to DSI. These comments included their opinion that:

- open access is important for research, especially for non-commercial research;
- open access has contributed to rapid developments in science;
- access to data should not be hampered by legal restrictions or administrative procedures/bureaucracy;
- restrictions on open access to DSI may slow down the exchange of data, slow scientific progress or potentially stop some types of research altogether, e.g. genomics research; and
- research in all countries would potentially be affected by restrictions on open access.

For several of the respondents, their concerns about possible restrictions on open access to DSI seem to be based on their experiences with obtaining genetic resources according to bilateral ABS rules under the Nagoya Protocol. These respondents indicated strong concerns that similar procedures may somehow apply to access to DSI.

4.5 Experiences with the Nagoya Protocol and PIC/MAT

4.5.1 Familiarity with the Nagoya Protocol

Almost two thirds of the respondents indicated that they were familiar with the term "access and benefit sharing" under the Nagoya Protocol. Of those, 81 respondents claimed to have had concrete experiences with obtaining PIC and MAT for genetic resources. However, this figure should be treated with some caution as it is possible that some respondents understood PIC and MAT to be a memorandum of understanding, a material transfer agreement or another type of research agreement which is entered into with a research institute in a country providing genetic resources rather than PIC and MAT, which are obtained from and negotiated with the relevant authorities and other parties, e.g. local communities. There is some support for assuming that the terms PIC and MAT were not correctly understood by all respondents. Two respondents, for example, indicated that they had never heard of ABS under the Nagoya Protocol but had always obtained PIC and MAT. These responses clearly contradict one another.

Assuming that the respondents really have had concrete experience with obtaining PIC and MAT, there still seems to be a gap between the number of people who claim to be familiar with the term ABS and those who have obtained PIC and MAT. The reason for this gap is not clear. A few comments made by respondents indicated that they work with genetic material which does not fall within the scope of the Nagoya Protocol, e.g. meaning that they work with material that was collected before the Nagoya Protocol came into force and therefore have not been affected by ABS rules. Comments such as this may point to a lack of understanding among researchers that the ABS regime has in fact been in place since the CBD came into force in 1993 and that ABS rules were developed and implemented in various countries, e.g. Brazil, long before the Nagoya Protocol came into force. This means that genetic material accessed prior to the Nagoya Protocol may nonetheless have been subject to ABS rules and the obligation to get PIC and MAT.

The survey asked specifically for the respondents' experiences with the Nagoya Protocol insofar as they relate specifically to DSI. Nevertheless, throughout the survey, a number of respondents took the opportunity to report on their general experiences with the Nagoya Protocol. Analysis of these comments indicates that the reported experiences were negative. It cannot be ruled out that respondents have had positive experiences with the Nagoya Protocol. However, if they have had positive experiences, these were not reflected in any of the comments made. Some of the themes that emerged included:

- uncertainty. This included uncertainty about the legal implications of sharing DSI, legal uncertainty due to the complexity of different ABS rules in different countries, and uncertainty about the future and how their work will be affected;
- bureaucracy and strict regulation. Respondents who have experience with the Nagoya Protocol regard it as a bureaucratic burden which creates a major drain on their resources, including both time and money. Some respondents indicated that it

was very complicated or difficult to get PIC and MAT. According to some respondents, this was largely due to a lack of the necessary administrative procedures and institutions in the countries providing genetic resources. There was also concern that it was difficult to obtain PIC and MAT due to very strict regulations in some countries; and

- negative impacts on their work including having to stop research, loss of collaborations with other scientists and restrictions on the ability to share data with former collaborators.

Although these comments were specifically addressed to the Nagoya Protocol, this again possibly ignores the fact that the ABS regime and the related issues may already have existed prior to the Nagoya Protocol coming into force in 2012.

4.5.2 Experiences with PIC and MAT and DSI

From the comments, it became apparent that there have been several cases where people have unsuccessfully attempted to get PIC and MAT to sequence genetic resources (n=4). Reasons provided by respondents for failing to get PIC and MAT included inadequate administrative procedures in the provider country, the application process being too complicated, long delays, and a perceived lack of understanding among the relevant officials regarding the respondents' scientific objectives and the purpose for which DSI was to be generated and used.

Of the respondents who claimed that they have obtained PIC and MAT, the majority (n=76) indicated that sequencing had been explicitly referred to in those documents. In roughly two thirds of those cases, sequencing had always been permitted, as shown in Figure 14. As mentioned above, it needs to be kept in mind that these respondents may be referring to memorandum of understanding or similar agreement with a research institute and not PIC and MAT per se.

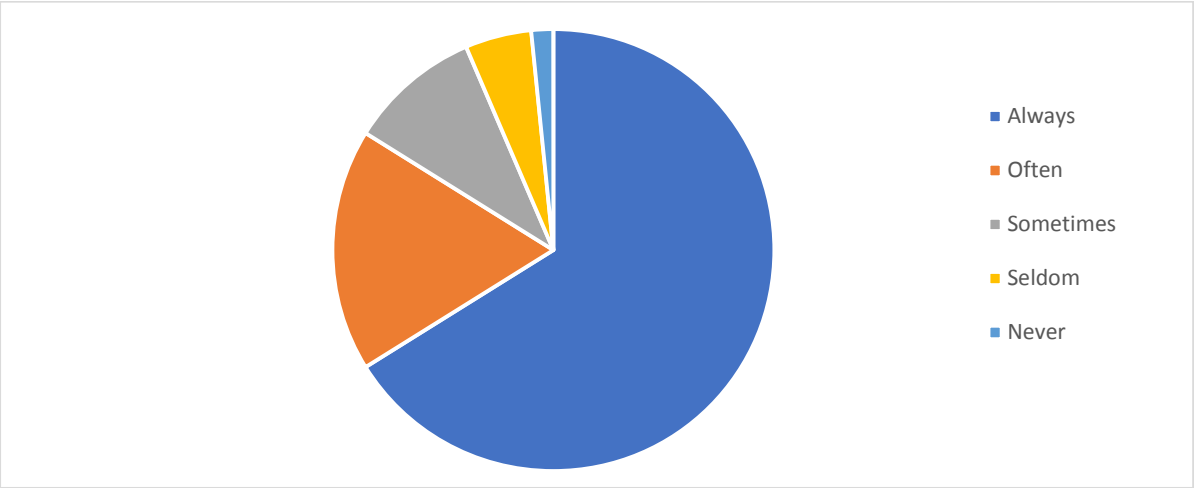

Figure 14 Frequency with which sequencing of genetic resources was explicitly permitted.

Those respondents who had not been permitted to sequence genetic material were asked to provide the reasons why they were not granted permission. Several respondents (n=5) reported that they were not provided with any reasons or explanation from the relevant authorities as to why permission was refused. In the opinion of other respondents, sequencing was not permitted:

- to prevent screening for natural compounds and thus prevent potential commercial uses (n=3);
- due to country policy, e.g. sequencing should be conducted in the country providing the genetic resources (n=2);
- due to bureaucratic problems (n=2); and
- due to intellectual property rights over the material (n=2).

It is not clear from the comments made by these respondents whether any of these reasons were made explicit, e.g. in the course of negotiations or in writing, and therefore should be treated with some caution because this may only be an assumption that was made by these respondents.

Those respondents who were unable to get PIC and MAT for sequencing or who received PIC and MAT without permission to sequence reported on various consequences of this. These included:

- termination of the project/abandonment of the research question;
- being unable to conduct the project according to plan;
- obtaining permits for sequencing at a later stage of the project;
- having limited scientific insight into the genetic resources;
- the material could not be used / the specimen became useless;
- sequences generated at the request of the provider country (by another government department) could not be used;
- the work was obsolete;
- wasted resources in terms of time spent negotiating contracts;
- wasted time and funding/wages for scientific work which ultimately could not be used;
- having to find alternatives;
- doing other research; and
- having no interest to continue working with the same partners and having plans to change countries/resources in future.

Several respondents also indicated that there were cases where PIC and MAT did not permit them to publish sequences. Again, a couple of these respondents were not provided with reasons as to why publishing was not permitted. Several others indicated that publication and sharing was prohibited due to intellectual property.

The respondents who were not permitted to share or publish their DSI reported on various consequences of this, including:

- having to exclude certain resources from their research, i.e. they could not use the material;
- being unable to include the data in their results;
- the quality of the research being affected;
- having to stop their research;
- there being limitations on cooperation or collaboration with third parties;
- having no intention to work in the same country again; and
- it being unlikely that they get further funding for the research.

4.5.3 Potential consequences if sequencing or publication of DSI is not permitted in future

Respondents were also asked a hypothetical question regarding the potential consequences for them if in future, PIC and MAT would not permit sequencing of genetic resources or the sequenced data to be shared/published. Given that this was a hypothetical question, the answers do need to be treated with some caution as this is only what the respondents think might happen. Of the 222 respondents who responded to this question, only one person indicated that this would not be a problem. The remaining respondents saw the potential consequences negatively.

A number of respondents used strong, negative adjectives and nouns to describe the potential consequences, including bad, terrible, disaster, catastrophe and devastating. A number of respondents also indicated that it would make it impossible for them to continue with their work or more generally, it would make research in their field impossible. Respondents provided examples of the affected fields, including evolutionary biology, taxonomy, metabolic engineering and biotechnology. Some respondents indicated that their research would be still possible but it would be a severe detriment and would slow both their own progress and the progress of their field in general, e.g. by necessitating a return to classical methods (where possible). Respondents also suggested that there would be implications for resources as research would become much more time consuming and much more expensive.

A major problem for respondents would be the inability to publish. Respondents indicated that publishing DSI is critical for their work and that the inability to publish would result in a loss of information available to the wider scientific community and also potentially the loss of research funding.

Some respondents indicated that they would change their field or the focus of their research. Others indicated that this would potentially have a major impact on their careers, that they may lose their employment or would prefer to enter early retirement than to continue with their work.

Some respondents expressed concern that restrictions on access to DSI may result in avoidance of certain countries and thus undermine research and conservation efforts in those countries. A couple of respondents also suggested that restrictions on access to DSI could impact on research benefiting human health.

The possible loss of collaboration was also raised. Some respondents indicated that they would no longer want to collaborate with partners in countries with restrictive approaches to access and they would search for collaborations in countries with less restrictive approaches.

Some respondents indicated that they would restrict their work to Germany and other EU countries to avoid such restrictions (although some EU countries also regulate access to genetic resources) or they would avoid genetic resources which are subject to restrictions on sequencing and publishing. A couple of respondents suggested that avoidance of certain regions and genetic resources etc. could potentially have implications for the quality of their research.

4.6 Access and Benefit Sharing for DSI

The survey also attempted to gauge an initial response towards potential ABS arrangements in the context of DSI.

4.6.1 Access

There were very mixed responses towards the statement "It is fair to give something in exchange for having access to digital sequence information/data", as shown in **Fehler! Verweisquelle konnte nicht gefunden werden.**. Almost one third of respondents disagreed strongly with this statement, which is probably unsurprising given the importance the respondents placed on having free, easy and unrestricted access to DSI.

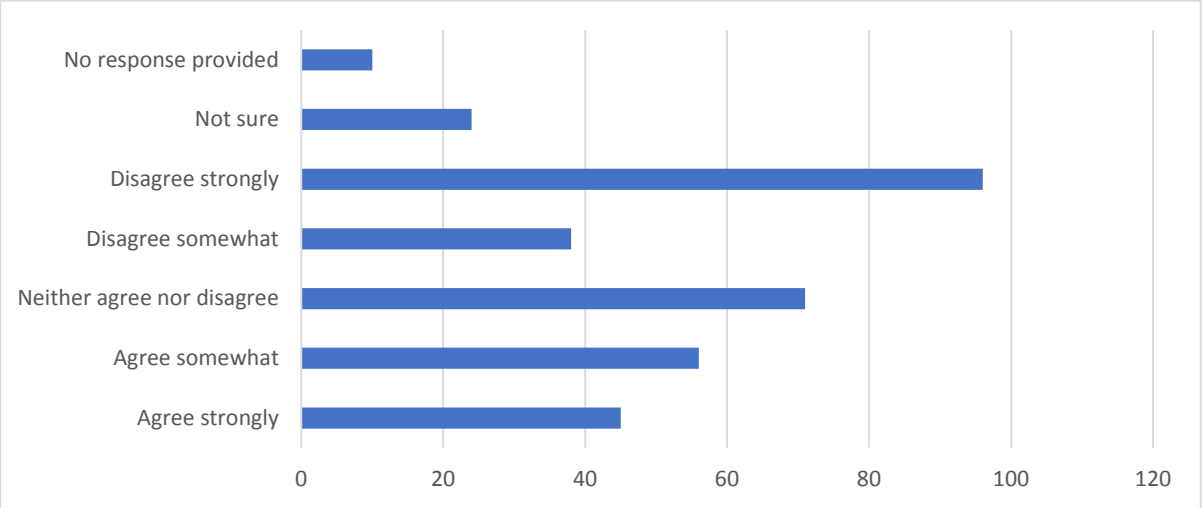

Figure 15 Level of agreement with the statement "It is fair to give something in exchange for having access to digital sequence information/data".

Most respondents disagreed in some way with the idea of paying for access to DSI, as shown in Figure 16. Nevertheless, there also seems to be a small group of respondents who show some willingness to pay for access to data.

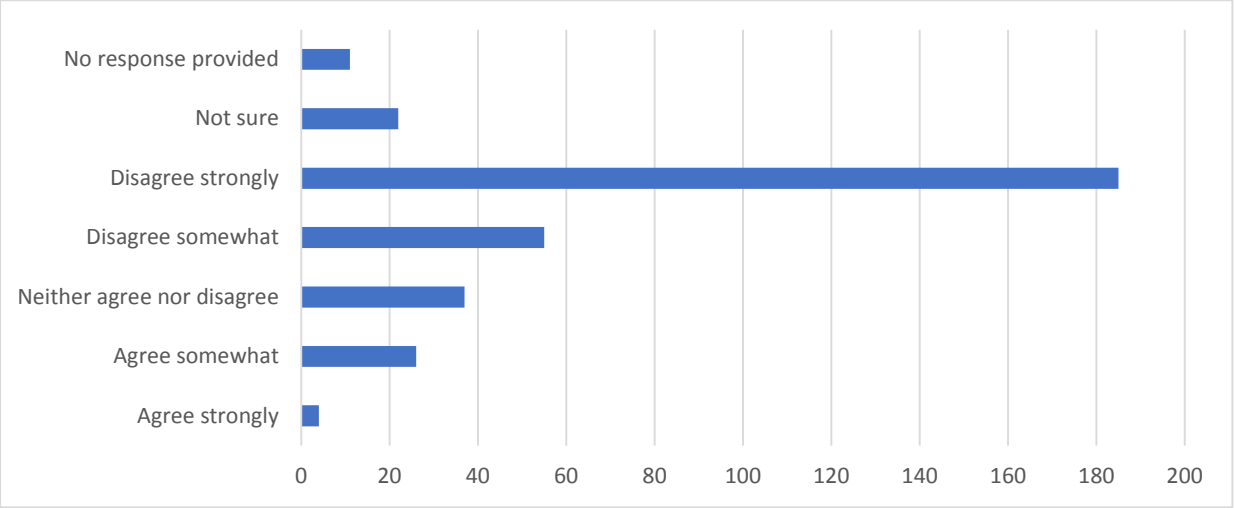

Figure 16 Level of agreement with the statement "I would pay in exchange for being able to access and use digital sequence information/data".

However, most respondents disagreed strongly with the idea of paying for access each time (n= 266) they access DSI and also having to pay for access over set period of time (e.g. one year) (n=225).

Based on comments made by the respondents, several reasons for this resistance among the respondents to the idea of linking access to DSI to payments emerged, including:

- open access is critical, especially for non-commercial research;
- access to DSI, particularly in public databases, should be freely accessible;
- respondents do not have funding to pay for access to DSI;
- there is no commercial value in the DSI they work with;
- making access to DSI fee-based would potentially restrict research to well-funded institutions that could afford to pay;
- fee-based access to DSI would be especially disadvantageous for researchers in poorer countries; and
- a payment system would further increase bureaucracy.

4.6.2 Benefit sharing

In their comments, some respondents indicated that they are already sharing benefits with countries which provide genetic resources, e.g. through the transfer of knowledge, collaborations and joint authorship of papers. Some respondents also suggested that they are sharing benefits by publishing their data and the results of their work.

Most respondents (n=231) disagreed strongly or somewhat with the statement that "countries that provide genetic resources should get something when digital sequence information/data from those genetic resources are used for basic research". On the other

hand, there was some support for the idea of benefits from commercialization of products based on DSI flowing back to countries which provide the genetic resources. Most respondents agreed strongly or somewhat (n=248) with the statement that "countries that provide genetic resources should get something when digital sequence information/data from those genetic resources are used for commercial purposes". This high level of agreement is probably unsurprising considering most of the respondents are engaged in non-commercial research and therefore would be unaffected by any potential benefit sharing obligations arising from the commercial use of DSI.

Respondents made various comments regarding the potential commercial value of DSI. These included that:

- there is a misconception regarding the potential monetary value of DSI;
- a lot of the data available in public repositories do not have commercial value, e.g. many sequences are used for phylogenetic research;
- there is a need for a clear distinction to be made between the different types of sequences and their potential value;
- sequences used for commercial purposes are typically exploited prior to publication or they are not published;
- some sequences may be linked to the production of chemical compounds, which could potentially have future application;
- potential value is unpredictable and often unknown when research begins.

This study did not attempt to explore the value of DSI or the validity of the comments made. However, the respondents' comments highlight the need for a better understanding of the value of DSI.

Other comments made by the respondents with respect to potential benefit sharing based on DSI included that:

- data sharing/publication should be standard;
- open access to DSI is a benefit, which could potentially be shared by all countries;
- countries that do not have the capacity to generate DSI themselves may especially benefit from open access;
- there are potentially non-monetary benefits associated with open access to DSI;
- public databases are often under multinational control and somewhat "fragile", meaning that it is important that their running costs are outweighed by the benefit of making the data available to the wider community;
- benefit sharing would be unclear as there is a great deal of overlap between sequences and therefore a sequence could possibly be obtained from multiple sources;
- there is a need for clear definition between commercial and non-commercial research with respect to benefit sharing obligations;

- a focus on monetary benefits from DSI would ignore the investment and effort made by researchers who generate DSI and make it available for the whole research community; and
- in the case of monetary benefit sharing, it is unclear who would be eligible to receive monetary benefits and whether the use of these funds would be transparent.

These comments indicate the need for a better understanding of the value of open access to DSI for the relevant stakeholders.

One respondent raised the idea of a multilateral benefit sharing system for DSI but suggested that such a mechanism would face major challenges due to the large volumes of data being exchanged on a daily basis and the potential risks associated with imposing conditions on data transactions.

4.7 The term "digital sequence information on genetic resources"

"Digital sequence information on genetic resources" is the term used in the COP/COP-MOP decisions but as discussed above, no definition has been provided for this term yet. The survey instrument attempted to find out what the respondents thought of the use of the terms "data" and "information".

Approximately 40% of the respondents indicated that they do not see any difference between "information" and "data", which suggests that using the term interchangeably would not be problematic, at least in their view. Some comments made by respondents suggested that trying to separate "data" and "information" would be impossible or "absurd".

Over half the participants, however, indicated that there is a difference between "data" and "information". The respondents were then asked to explain this difference.

Respondents suggested that "data" could be understood as:

- the plain sequence;
- the exact sequence of nucleotides;
- a string or collection of letters;
- something that is raw;
- sequences which have not been annotated;
- the result or output of the sequencing process; and
- something that does not involve any intellectual output.

One respondent suggested that although ways of interpreting data change, the data stay the same.

A couple of respondents took an alternative view of the term “data”, stating that the nucleic acids or the organism could be regarded as data and information as the output of the sequencing process.

Respondents suggested that “information”:

- is derived or extracted from data;
- is something which is generated once a raw data set has been processed and filtered;
- refers to annotated sequences;
- includes sequences with metadata;
- is the metadata;
- is not found in non-coding parts of a sequence;
- involves analysis or interpretation;
- is the output of the human mind;
- requires (considerable) expertise for its production;
- depends on the research question;
- has different levels and can be general;
- does not necessarily have to include the nucleotide sequence itself;
- conveys meaning; and
- is useful.

There were also couple of contradictory views, e.g. that data comes from information and that information represents the letters of the nucleotide sequence.

Almost 70% of the respondents agreed strongly or agreed somewhat with the statement that “Digital sequence data have to be interpreted before they have value”. There were mixed responses to the statement that “Digital sequence information is not more valuable than digital sequence data”, as shown in Figure 17, with a little more than a half disagreeing in some way with this statement. This suggests that for some respondents, interpretation of data may lead to an increase in value of the data in some way and assuming that information is the product of that interpretation, information could be regarded as having more value.

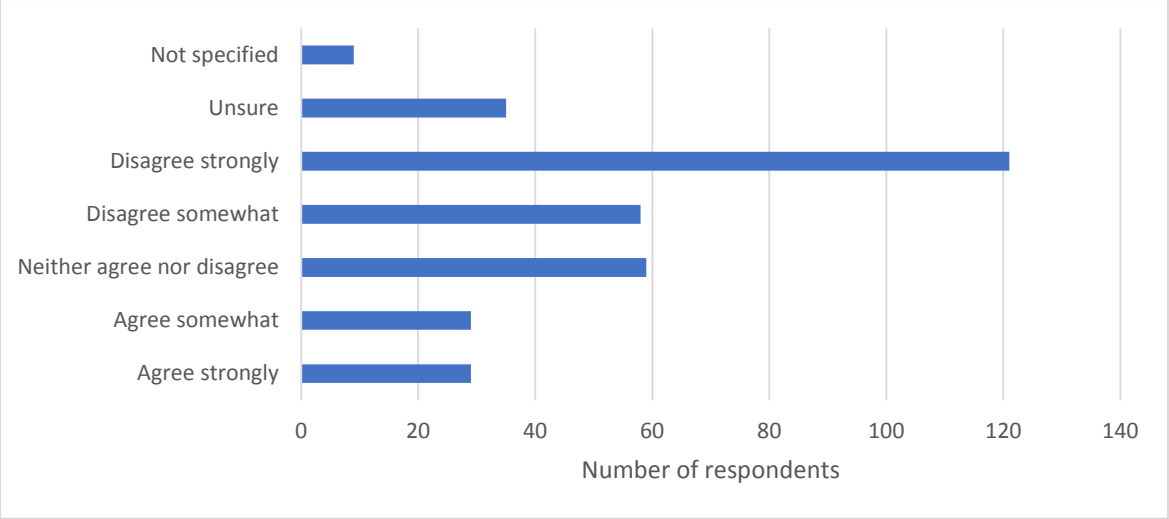

Figure 17 Level of agreement with the statement "Digital sequence information is not more valuable than digital sequence data".

Although a number of respondents see that there is a difference between data and information, identifying a point where data ends and information begins, i.e. a so-called "cut-off" point, was more challenging. Some participants indicated that they had no idea how to define this or that it was very difficult (if not impossible), e.g. because creating information is context dependent and depends upon the relevant research question. This suggests that identification of an arbitrary cut-off between data and information could be problematic.

Many people simply referred again to the difference between data and information. For many of those people who regarded data as the raw sequence of letters, they thought that anything above and beyond the letters of the sequence amounts to information, meaning that information is created as soon as data processing begins. Others suggested that processed or curated sequences are still data. Some suggested that the cut-off point between data and information may be distinguished based on whether there are metadata or annotations. However, there were also differences of opinion about which type of annotation would be necessary, for example, whether this would include annotation using computer-based algorithms or whether the annotation would have to be done by a scientist.

Other suggested cut-off points, i.e. when data becomes information included:

- when an analysis or experiment begins;
- analysis plus interpretation; or
- the identification of function.

The diverging opinions of the respondents demonstrate the difficulties associated with using arbitrary and undefined terms. Although the terms data and information have ordinary meaning, it appears that understanding of this meaning is not universal, even to the

respondents who use these terms regularly. This suggests that clear definitions are needed but at the same time, creating definitions that make sense in practice could be very difficult.

5. Discussion

5.1 The use of DSI in Germany

This exploratory study was intended to shed some light on the stakeholders in Germany who work with DSI. The respondents who participated in the study form a heterogeneous group, coming from a range of disciplines in the life sciences and related fields. The respondents use sequences from a variety of different organisms from all over the world and for various purposes. Most of the participants generate their own DSI and obtain DSI from other sources, e.g. through sharing or public databases such as GenBank. The study showed that DSI generated through their own work and DSI obtained from other sources are both important for most of the participating researchers. Most of these researchers also expect to use DSI in the future.

Although the results of this study suggest that DSI has become a very important aspect of their work and research, it does not necessarily mean that the participating researchers are able to work exclusively with DSI, especially with DSI obtained from other sources. Most of the participating researchers still generate their own data, i.e. they sequence genetic resources themselves, and for the majority, the nature of their research/work also appears to make access to physical genetic resources necessary. Few respondents indicated that they work with genetic resources that have been created synthetically and most respondents do not expect to be able to work exclusively with synthetic genetic resources in the future.

As most of the respondents share or publish at least some of the DSI generated through their work, it might be assumed that at least some form of non-monetary benefit sharing is taking place. Nevertheless, it certainly does not appear that all DSI generated through their work finds its way into public databases etc. and thus becomes available to the whole scientific community. This points to the need for a better understanding of the data sharing practices of the research community and how these practices are governed.

The respondents highlighted the importance having open access to DSI, especially for basic research. There seem to be strong concerns amongst these researchers that access to DSI could somehow become more restrictive and that this would impact on their work.

Although distinguishing between commercial and non-commercial research may be important, e.g. for determining different benefit sharing obligations, the results of the survey seem to confirm that the line between basic and commercial research is not always clearly defined. There are a wide range of uses for DSI in Germany, and for most of the

people who participated in this study, these uses are not typically regarded as being commercial in nature, e.g. taxonomic work, phylogenetic analyses etc. Nevertheless, it cannot be completely ruled out that the knowledge created in basic research does not become part of a wider body of knowledge that somehow underpins commercial research and product development. Although respondents indicated that they have little way of knowing whether their published DSI is accessed and how it is subsequently used, many still claimed to know that their published DSI has been reused in some way by others and in a few cases, this published DSI had been used for the development of products.

Whereas most of the participants in this study indicated that their work is not commercial in nature, a few reported that they are using DSI for product development. Despite evidence that there is some use of DSI for commercial purposes in Germany, it is not clear whether any commercial benefits are already accruing to certain actors as a consequence of this use or whether future commercial benefits would potentially be attributable to this use.

A few respondents reported on concrete examples of not being able to get PIC and MAT to sequence genetic resources and/or to publish their DSI. In a couple of cases, this may have been attributable to the fear of provider states of potential commercialization without consent. In the view of the researchers, the inability to get permission to sequence genetic resources and/or publish their DSI had negative impacts on their research, projects and collaborations. Most respondents also indicated that if hypothetically, PIC and MAT would not allow sequencing or the publication of DSI, there would be negative consequences for their work.

This study suggests that any changes to access arrangements for DSI or any restrictions on sequencing of genetic resources and/or publication of the resulting DSI could potentially affect a wide number of individual researchers, institutions and industry actors, and in the case of the people involved in this study, most of those affected would be researchers employed in the public sector who are involved in non-commercial research and may potentially be contributing to the conservation and sustainable use of biodiversity through their research. There are very strong concerns amongst the respondents that any restrictions in future could potentially hinder their work, make it more costly or even make it impossible.

5.2 The implications for Nagoya Protocol from the use of DSI

There is a need for further clarification of the term “DSI” and its scope, i.e. which types of sequences are covered by the discussion on DSI and whether a distinction between data and information is relevant or necessary. This study indicates that among the participating researchers, there are diverging understandings of these information concepts, which suggests that it could be quite challenging to create workable definitions that make sense in this context.

The participants in this study sometimes work with large numbers of sequences, data are shared privately amongst researchers or accessed from public databases, and in some cases, the source of the original genetic resources is unknown. Once published, the researchers themselves have no control over this DSI and little way of knowing about who accesses it and how it is subsequently used. This suggests that if the Nagoya Protocol applies for DSI, there could be some major challenges for the access, benefit sharing and compliance pillars of the protocol.

As most of the respondents generate their own DSI, they presumably also need access to physical specimens of genetic resources for this purpose. In cases where they need genetic resources from parties to the Nagoya Protocol, they will have to comply with any relevant legislation in that state, with any obligations arising from PIC and MAT as well as with the due diligence obligations under the EU Regulations for any genetic resources falling within the scope of the regulations.

Most of the researchers in this study work with published DSI but only a relatively small number of participants indicated that they can already work exclusively with DSI from other sources, i.e. they have no need to access genetic resources and sequence them. This suggests that there has not been a complete "dematerialization" of research and development yet. For some of the respondents, however, the limiting factor is the availability of data in publicly accessible databases. As such, it cannot be entirely ruled out that as more information/data become available and technology improves, more people will be able to work exclusively with published DSI in the future. On the other hand, there may equally be some aspects of research and development which will not be possible without using physical genetic resources in combination with DSI.

This study did not attempt to identify or quantify benefits arising from the use of DSI and given that the discussions at the international level around DSI are at a very early stage, statements about potential access arrangements and benefit sharing arising from the use of DSI are quite premature. The results indicated that the participating researchers are in favour of open, freely available access to DSI and are against having to give something or pay something in exchange for this access. Although there seems to be some support for the idea of benefit sharing arising from the commercialization of products based on DSI, there was less support for the idea of benefit sharing arising from use of DSI for basic research. There were also a number of comments about the value of open access to the research community. This study did not explore the validity of statements about the value of open access to DSI for the global research community or the public generally. However, further analysis of any benefits arising from open access to DSI, how these benefits are generated and by which actors as well as who benefits could potentially provide much more insight into these issues.

5.3 Limitations of the study

This was an exploratory study. The online survey attempted to gain insight into many different aspects of the work by the participants. This led to the survey instrument becoming quite long, which could mean that participants were somewhat fatigued by the end of the survey, that the length of the survey discouraged people from participating or that people were discouraged from completing all questions. Nevertheless, many participants completed the whole survey and took the time to provide answers to open questions, which suggests that there is substantial interest in the issue.

An online survey was chosen to reach a maximum amount of people with the highest level of convenience. While providing a broad perspective on the use of DSI, it does not offer the same level of insight as expert interviews, for example. As the study was conducted in online form, it was also not possible to assist those participants who may have misunderstood questions or were unclear about the meaning of certain terms, such as "synthetic genetic resources" etc. The survey instrument could have been improved by providing definitions (where such definitions exist), although adding definitions to the survey would also have increased the length of the survey instrument even further.

Although the study was not intended to be representative, it would have been preferable if more participants from industry and commercial research had participated. It is not clear why the levels of participation were low from these sectors, despite key industry associations being contacted.

The study was conducted in English in order to make it accessible to the wider international community and also so that foreign scientists working in Germany who cannot speak German could participate. As most people working in research need to read scientific papers in English, publish in English and present at international conferences etc., it was assumed that participants would not have problems completing the survey in the English language. A couple of people chose to respond to open questions in German and these responses were included in the dataset. It is possible that the formulation of certain questions may have required further clarification, e.g. the difference between past concrete experiences and a hypothetical situation. It appears that this grammatical distinction was not understood properly by several participants based on their responses, e.g. because the response did not fit with the question asked or contradicted a previous response. This meant that in some cases, a few responses had to be excluded from the analysis of specific questions. In other cases, the responses were formulated in such a way that the meaning was not clear. Nevertheless, almost all respondents seemed to have no problems with the language and it can be assumed that it was not a major disadvantage that the study was conducted in English.

6. Outlook

This study has provided some insight into the stakeholder landscape in Germany with respect to DSI and its use, especially into use of DSI for non-commercial research purposes. Nevertheless, there are a number of questions which require further consideration. The low levels of participation by commercial researchers and industry limit the understanding of the use of DSI by these stakeholder groups. At least based on the responses of those few respondents who participated in this study, it seems that DSI would be important for these stakeholders and this suggests that there may be a need to investigate this further. Although the study showed that DSI is important to the individual researchers who participated in the study, there was no attempt to define or to quantify the value of DSI to the researchers in any way. Although researchers in Germany use DSI, including for the development of products, it is not clear to what extent actual benefits, either monetary or non-monetary, are generated through this use. A clearer understanding of these benefits as well as benefits arising from open access would help further. Based on how DSI is accessed, used and shared by the participants, it seems likely that applying the existing bilateral ABS system to DSI would be problematic for many of these stakeholders. This indicates the need for further consideration of the potential impacts as well as any alternative arrangements that could potentially meet the needs of the stakeholders involved should ABS obligations apply.

7. References

Allan, E. 2014. Metagenomics. Virulence. 5(3), 397-398. DOI: 10.4161/viru.28057.

Allendorf, F.W., Hohenlohe, P.A. and Luikart, G. 2010. Genomics and the future of conservation genetics. Nature Reviews Genetics. 11, 697-710.

Amorim, A. 2013. Population Genetics. Brenner's Encyclopedia of Genetics. 2nd edition. 5, pp. 407-411. doi:10.1016/B978-0-12-374984-0.01195-5

Bagley, M.A. 2017. Towering Wave or Tempest in a Teapot? Synthetic Biology, Access & Benefit Sharing, and Economic Development, in S. Frankel and D. Gervais (eds.). The Internet and Intellectual Property: The Nexus with Human and Economic Development. (Victoria University Press forthcoming).

Bagley, M.A. 2015. Digital DNA: the Nagoya Protocol, intellectual property treaties, and synthetic biology. Wilson Center.

Bagley, M.A. and Rai, A. 2013. The Nagoya Protocol and Synthetic Biology Research: A Look at the Potential Impacts. Wilson Center.

Baškarada, S. and Koronios, A. 2013. Data, Information, Knowledge, Wisdom (DIKW): A Semiotic Theoretical and Empirical Exploration of the Hierarchy and its Quality Dimension. Australasian Journal of Information Systems. 18(1), 5-24.

Brocchieri, L. 2001. Phylogenetic Inferences from Molecular Sequences: Review and Critique. Theoretical Population Biology. 59, pp. 27-40.

Bygrave, L.A. 2015. Information Concepts in Law: Generic Dreams and Definitional Daylight. Oxford Journal of Legal Studies, 35(1), 91–120. doi:10.1093/ojls/gqu011.

CBD Secretariat. 2018. Synthesis of views and information on the potential implications of the use of digital sequence information on genetic resources for the three objectives of the Convention and the objective of the Nagoya Protocol. CBD/DSI/AHTEG/2018/1/2. CBD Secretariat, Montreal.

Charleston, M. 2013. Phylogenetics. Brenner's Encyclopedia of Genetics, 2nd edition. 5, pp. 324-325. doi:10.1016/B978-0-12-374984-0.01160-8.

Coorssen, J.R. 2013. Proteomics. Brenner's Encyclopedia of Genetics, 2nd edition. 5, pp. 508-510. doi:10.1016/B978-0-12-374984-0.01231-6

Culligan, E.P., Sleator, R.D., Marchesi, J.R. and Hill, C. 2014. Metagenomics and novel gene discovery. Virulence. 5(3), 399-412. DOI: 10.4161/viru.27208.

De Jonge, B. 2013. Towards a Fair and Equitable ABS Regime: Is Nagoya Leading us in the Right Direction? Law, Environment and Development Journal. 9(2), 241-255. http://www.lead-journal.org/content/13241.pdf

De Lorenzo, V. and Danchin, A. 2008. Synthetic biology: discovering new worlds and new words. EMBO reports. 9 (9), 822-827.

Fecher, B., Friesike, S. and Hebing, M. 2015. What Drives Academic Data Sharing?. PLoS ONE. 10(2). e0118053. doi:10.1371/journal.pone.0118053.

European Commission. 2016. Guidance document on the scope of application and core obligations of Regulation (EU) No 511/2014 of the European Parliament and of the Council on the compliance measures for users from the Nagoya Protocol on Access to Genetic Resources and the Fair and Equitable Sharing of Benefits Arising from their Utilisation in the Union. Commission Notice 2016/C 313/01.

Glowka, L., Burhenne-Guilmin, F. and Synge, S. in collaboration with McNeely, J.A. and Gündling, L. 1994. A Guide to the Convention on Biological Diversity, IUCN, Gland and Cambridge.

Greiber, T., Peña Moreno, S., Åhrén, A., Nieto Carrasco, J., Kamau, E.C., Medaglia, J.C., Oliva, M.J., Perron-Welch, F. in cooperation with Ali, N. and Williams, C. 2012. An Explanatory Guide to the Nagoya Protocol on Access and Benefit-sharing. IUCN, Gland, Switzerland.

Hood, L. and Galas, D. 2003. The digital code of DNA. Nature. 444-448.

Huttenhower, C. and Hofmann, O. 2010. A Quick Guide to Large-Scale Genomic Data Mining. PLoS Comput Biol. 6(5). e1000779. doi:10.1371/journal.pcbi.1000779

Idle, J.R. and Gonzalez, F.J. 2007. Metabolomics. Cell Metabolism. 6, 348-351. DOI 10.1016/j.cmet.2007.10.005

IISD Reporting Services. 2016. Summary of the UN Biodiversity Conference: 2-17 December 2016. Earth Negotiations Bulletin. 9 (678). Online at: http://www.iisd.ca/biodiv/cop13/enb/

Jaspars, M. 2017. Categories of information and types of data incorporating different levels of processing and analysis, Presentation at IUCN Workshop entitled “Exchange of views on building a consensus on benefit sharing” at New York University Law School, New York, USA, 1 April 2017. Cited in Laird, S. and Wynberg, R. with contributions from Arash Iranzadeh and Anna Sliva Kooser. 2018. A Fact-Finding and Scoping Study on Digital Sequence Information on Genetic Resources in the Context of the Convention on Biological Diversity and the Nagoya Protocol. CBD/DSI/AHTEG/2018/1/3. CBD Secretariat, Montreal.

Kaur, S. 2013. Genomics. Brenner’s Encyclopedia of Genetics, 2nd Edition, 3, 310-312. doi:10.1016/B978-0-12-374984-0.00642-2

Koenig, H., Dorado-Morales, P. and Porcar, M. 2015. Responsibility and intellectual property in synthetic biology. A proposal for using Responsible Research and Innovation as a basic framework for intellectual property decisions in synthetic biology. EMBO reports. 16 (9).

Laird, S. and Wynberg, R. 2016. Locating Responsible Research and Innovation within Access and Benefit Sharing Spaces of the Convention on Biological Diversity: the Challenge of Emerging Technologies. Nanoethics. DOI 10.1007/s11569-016-0268-z.

Laird, S. and Wynberg, R. with contributions from Arash Iranzadeh and Anna Sliva Kooser. 2018. A Fact-Finding and Scoping Study on Digital Sequence Information on Genetic Resources in the Context of the Convention on Biological Diversity and the Nagoya Protocol. CBD/DSI/AHTEG/2018/1/3. CBD Secretariat, Montreal.

Lawson, C. and Rourke, M. 2016. Open Access DNA, RNA and Amino Acid Sequences: The Consequences and Solutions for the International Regulation of Access and Benefit Sharing. Griffith Law School Research Paper No. 16-12.

Lucchi, N., (2013). Understanding genetic information as a commons: from bioprospecting to personalized medicine. International Journal of the Commons. 7(2), pp. 313–338.

Mannheim, B. 2016. Regulation of synthetic biology under the Nagoya Protocol. Nature Biotechnology. 34(11), 1105-1105.

Manzella, D. 2016. The Global Information System and genomic information: transparency of rights and obligations. Background study paper n. 10. Item 6 of the Provisional Agenda. International Treaty on Plant Genetic Resources for Food and Agriculture First Meeting of the Scientific Advisory Committee on the Global Information System of Article 17 of the Treaty. FAO.

Marx, V. 2013. The challenges of big data. Nature. 498, 255.

McGettigan, P.A. 2013. Current Transcriptomics in the RNA-seq era. Opinion in Chemical Biology. 17, 4–11.

Medaglia, J.C. 2015. Access and Benefit-Sharing: North–South Challenges in Implementing the Convention on Biological Diversity and its Nagoya Protocol in Alam, Shawkat, Atapattu, Sumudu, González, Carmen, Razzaque, Jona (eds). in “International Environmental Law and the Global South”, Cambridge University Press.

Morgera, E., Buck, M. and Tsioumani, E. 2014. Unraveling the Nagoya Protocol. A Commentary on the Nagoya Protocol on Access and Benefit-sharing to the Convention on Biological Diversity. Koninklijke Brill. Leiden. The Netherlands.

Oldham, P., Hall, S. and Burton, G. 2012. Synthetic Biology: Mapping the Scientific Landscape. PLoS ONE 7(4). e34368. doi:10.1371/journal.pone.0034368.

Ramsden, J. 2015. Bioinformatics: An Introduction. Springer. London.

Richerzhagen, C. 2011. Effective governance of access and benefit-sharing under the Convention on Biological Diversity. Biodivers Conserv. 20, 2243–2261. DOI 10.1007/s10531-011-0086-0

Rosendal, K. and Andresen, S. 2016. Realizing access and benefit sharing from use of genetic resources between diverging international regimes: the scope for leadership. Int Environ Agreements. 16, 579–596. DOI 10.1007/s10784-014-9271-4.

Schmidt B., Gemeinholzer, B. and Treloar, A. 2016. Open Data in Global Environmental Research: The Belmont Forum's Open Data Survey. PLoS ONE 11(1). e0146695. doi:10.1371/journal.pone.0146695.

Scott, D., Abdelhakim, D., Miranda, M., Höft, R. and Cooper, H.D. 2015. Potential positive and negative impacts of components, organisms and products resulting from synthetic biology techniques on the conservation and sustainable use of biodiversity, and associated social, economic and cultural considerations. Part I of: Synthetic biology. Secretariat of the Convention on Biological Diversity. Montreal, Technical Series No. 82, 60 pages.

Shokralla, S., Spall, J.L., Gibson, J. and Hajibabaei, M. 2012. Next-generation sequencing technologies for environmental DNA research. Molecular Ecology. 21, 1794–1805. doi: 10.1111/j.1365-294X.2012.05538.x

Spranger, T.M. 2017. Expert opinion on the applicability of the Convention on Biological Diversity and the Nagoya Protocol to digital sequence information. Commissioned by the German Federal Ministry of Education and Research, Berlin. Available at https://www.bmbf.de/files/Legal_opinion_DSI_Prof_Spranger_EN_BF.PDF

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. Wu, L., Read, E., Manoff, M. and Frame, M. 2011. Data Sharing by Scientists: Practices and Perceptions. PLoS ONE. 6(6). e21101.doi:10.1371/journal.pone.0021101.

Tenopir, C., Dalton, E.D., Allard, S., Frame, M., Pjesivac, I., Birch, B., Pollock, D. and Dorsett, K. 2015. Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide. PLoS ONE. 10(8). e0134826. doi:10.1371/journal.pone.0134826.

Tyler-Smith, C., Yang, H., Landweber, L.F., Dunham, I., Knoppers, B.M., Donnelly, P., Mardis9, E.R., Snyder, M. and McVean, G. 2015. Where Next for Genetics and Genomics? PLoS Biol 13(7). e1002216. doi:10.1371/journal.pbio.1002216.

Varshney R.K., Terauchi, R. and McCouch, S.R., (2014) Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. PLoS Biol. 12(6). e1001883. doi:10.1371/journal.pbio.

von Kries, C. and Winter, G. (2015). Defining commercial and non-commercial research and development under the Nagoya Protocol and in other contexts. In: E.C. Kamau, G. Winter and P-T. Stoll, eds., Research and development on genetic resources: public domain

approaches in implementing the Nagoya Protocol, 1st ed. Abingdon, Oxon [UK] ; New York, NY: Routledge.

Welch, E., Bagley, M., Kuiken, T. and Louafi, S. 2017. Potential implications of new synthetic biology and genomic research trajectories on the International Treaty for Plant Genetic Resources for Food and Agriculture (ITPGRFA or 'Treaty'). Scoping report prepared for ITPGRFA.

Yockey, H.P. 2005. Information theory, evolution, and the origin of life. Cambridge University Press, Cambridge.

Zimmermann, J., Abarca, N., Enk, N., Skibbe, O., Kusber, W-H. and Jahn, R. 2014. Taxonomic Reference Libraries for Environmental Barcoding: A Best Practice Example from Diatom Research. PLoS ONE. 9(9). e108793. https://doi.org/10.1371/journal.pone.0108793.

8. Annexes

Annex A: List of all disciplines/fields in which the respondents work or do research

Discipline/Field	No. of respondents
Agronomy (food, feed, fuel etc.)	13
Algal diversity	1
Animal behaviour	1
Animal breeding	6
Animal health	6
Biochemistry	42
Biocontrol (herbicides, insecticides, fungicides etc.)	8
Bioinformatics	67
Biology	168
Biomimetics	1
Biotechnology	48
Botany	2
Chemistry	9
Computer science	7
Developmental biology	3
Ecology	89
Engineering	2
Entomology	1
Evolutionary biology	148
Food	2
Genetic engineering	25
Genetics	91
Genomics	99
Health sciences	12
Identification of invasive and pest organisms	1
Immunology	1
Infection biology	1
Information science/technology	3
Integrative physiology	1
Medicine	14
Metabolomics	8
Metagenomics	27
Microbiology	90
Molecular biology	2

Discipline/Field	No. of respondents
Molecular systematics	1
Neurophysiology	1
Neuroscience	1
Pharmacy	5
Phenomics	1
Phylogenetics	2
Phylogenomics	1
Plant breeding	16
Plant pathology	9
Plant physiology	1
Plant science	1
Population genetics	1
Proteomics	9
Soil science	5
Synthetic biology	1
Taxonomy	111
Transcriptomics	47
Zoology	1
Not specified	3

<u>Annex B:</u> List of databases/repositories/platforms/tools etc. used by the respondents to publish the DSI generated through their own work or research

Database Name
Barley Draft Genome Explorer (BARLEX)
BioCatNet database system
Barcode of Life Database (BOLD)
BRaunschweig ENzyme DAtabase (BRENDA)
Comparative Genomics (CoGe)
Department of Energy Joint Genome Institute (DOE JGI)
Department of Energy Joint Genome Institute (DOE JGI) – Integrated Microbial Genomes (IMG)
DNA Databank of Japan (DDBJ)
Dryad Digital Repository (Dryad)
European Genome-phenome Archive (EGA)
European Bioinformatics Institute (EMBL-EBI) (generally)
EMBL-EBI – European Nucleotide Archive (ENA)
EMBL-EBI - Ensembl
EMBL-EBI – Short Read Archive (SRA)
Figshare
Forensic Science International
German Barcode of Life (GBOL)
Genome Announcements
International Barcode of Life (IBOL)
International Nucleotide Sequence Database Collaboration (insdc)
Mendeley Data
MG-RAST Metagenomics (metagenomics Rapid Annotation using Subsystem Technology)
National Centre for Biotechnology Information (NCBI) (generally)
NCBI – GenBank
NCBI - Short Read Archive (SRA)
NCBI - Protein Database
NCBI - Nucleotide
NCBI - Transcriptome Shotgun Assembly Sequence Database (TSA)
NCBI - PubMed
NCBI - Gene Expression Omnibus (GEO)
nhbs
National Institutes of Health (NIH)
Open Science Framework
OrthodDB
PANGEA

Database Name
Patent databases (unspecified)
Protein Data Bank (PDB)
Plant genomics and phenomics repository (PGP)
Plos ONE
Protist Ribosomal reference database (PR2 Database)
Q-Bank
Research Gate
Saccharomyces Genome Database (SGD)
Haplotype Polymorphism in Polyploid Wheats and their Diploid Ancestors (snpdb)
Sequence Read Archive
Short Read Archive
The Arabidopsis Information Resource (TAIR)
TreeBase

<u>Annex C:</u> List of databases/repositories/platforms/tools etc. used by the respondents to obtain DSI generated by others

Database Name
3DM
Arabidopsis Information Portal (Araport)
Aspergillus Genome Database (AspDG)
Barcode of Life Database (BOLD)
Barley Draft Genome Explorer (BARLEX)
BRaunschweig ENzyme DAtabase (BRENDA)
Broad Institute
CBS Database (Westerdijk Institute)
CELL
Comparative Genomics (CoGe)
CoryneRegNet
Centro di Ricera Interdipatimentale per le Biotechnologie Innovative (CRIBI)
CRAN-R
Department of Energy Joint Genome Institute (DOE JGI)
Department of Energy Joint Genome Institute (DOE JGI) – Integrated Microbial Genomes (IMG)
Department of Energy Joint Genome Institute (DOE JGI)- Genomes Online Database (GOLD)
Department of Energy Joint Genome Institute (DOE JGI)- MycoCosm
Department of Energy Joint Genome Institute (DOE JGI) - Plant Comparative Genomics (Phytozome)
DNA Databank of Japan (DDBJ)
Dictyoptera reference database (DictDB)
Dryad Digital Repository (Dryad)
EggNOG
EMBO
Encyclopedia of DNA Elements (ENCODE)
European Bioinformatics Institute (EMBL-EBI) (generally)
EMBL-EBI - European Nucleotide Archive (ENA)
EMBL-EBI – Ensembl
EMBL-EBI - Ensembl Plants
EMBL-EBI - UniProt
EMBL-EBI – Wormbase
EMBL-EBI – Sequence Read Archive (SRA)
EzTaxon
Feed Grains Database (grains)

Database Name
FlyBase
FungiDB
German Barcode of Life (GBOL)
Gene Ontology Consortium
GitHub
Google Scholar
Gramene Markers Database (Gramene)
Greengenes
Imicrobes
integrated microbial NGS platform (imngs)
International Nucleotide Sequence Data Collaboration (insdc)
Kyoto Encyclopedia of Genes and Genomes (KEGG)
Mendeley Data
Metagenomics RAST Server (MG-RAST)
MicrobesOnline
MycoBank
Nature
National Centre for Biotechnology Information (NCBI)
NCBI - CyanoBase
NCBI – GenBank
NCBI - Gene Expression Omnibus (GEO)
NCBI - Nucleotide
NCBI - Protein Database
NCBI – PubMed
NCBI - Reference Sequence Database (RefSeq)
NCBI - Sequence Read Archive (SRA)
NCBI - Database for Short Genetic Variations (SNPdb)
nhbs
OrthoDB
Protein ANalysis THrough Evolutionary Relationships (PANTHER)
Protein Data Bank (PDB)
Pathway Genome Databases (PGDBs)
Plant DNA C-values Database
Protist Ribosomal reference database (PR2 Database)
Q-Bank
Rapid Annotation using Subsystem Technology (RAST)
Research Gate
Ribosomal Database Project (RDP)
RIKEN

Database Name
Sanger Institute
SEED
Silva ribosomal database project (SILVA)
STRING
SwissProt
The Arabidopsis Information Resource (TAIR)
TIGR Maize Database (TIGR)
TreeBase
Tree of Life (ToL)
USCS Genome Browser

Annex D: Invitation and Survey Instrument

Survey questionnaire

Understanding the use of and experiences with digital sequence information in Germany

In 2018, **international negotiations** on the topic of **digital sequence information on genetic resources (DSI)** will take place between the parties to the **Convention on Biological Diversity** (CBD) and the **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 (Nagoya Protocol). Germany is a party to both of these international agreements.

We are investigating the **importance of DSI** for German stakeholders working in basic research, collections of genetic resources and industry. In particular, we are interested in understanding how DSI is generated, used and shared as well as any experiences stakeholders may have had so far with DSI in the context of access and benefit-sharing ("ABS") arrangements under the Nagoya Protocol.

As a **potential user of DSI**, you are invited to participate in our **study**, which has been made possible by the German Federal Agency for Nature Conservation (Bundesamt für Naturschutz, BfN) with funding from the German Federal Ministry for the Environment, Nature Conservation, Construction and Nuclear Safety (Bundesministerium für Umwelt, Naturschutz, Bau und Reaktorsicherheit, BMUB).

The results of the survey will be published in a report, which is intended to help the German government understand **key stakeholder positions** with regard to DSI and therefore inform its position in the upcoming international negotiations.

As a potential stakeholder, your participation in the survey would be very helpful.

The responses to this survey will be recorded **anonymously** and **cannot be linked** to you or your institution/employer in any way.

As the study is intended to inform international negotiations, it is being conducted in English. However, open questions in the survey may also be completed in German.

Your participation in this survey is voluntary. If you choose to participate, please complete the **online survey**, which should take about 15 **minutes, by 15 October 2018.**

[Link to survey]

If you have any questions about the study or the online survey, please contact Dr Axel Paulsch or Dr Cornelia Paulsch at the Institut für Biodiversität Netzwerk e.V. (info@biodiv.de or telephone +49 (0)941 381324 62).

Have you or has your organization/institute/association already formulated a **position paper** with respect to the issue of DSI, e.g. you have made a submission to the Secretariat of the CBD on this issue? These position papers can be included in the annex to the final report. If you wish to provide us with a copy of your position paper, please send it to info@biodiv.de or the following address:

Institut für Biodiversität
Nußbergerstr. 6a
93059 Regensburg

We thank you in advance for your participation in this study and reassure you once more that the information provided cannot be linked to you or your institution/employer in any way!

SURVEY ON THE USE OF AND EXPERIENCES WITH DIGITAL SEQUENCE INFORMATION/DATA

This survey is anonymous. No personally identifiable information is captured by the survey and itis not possible to trace your responses to you or your institution in any way.

Your responses will be combined with those of the other respondents and the results of the analysis will be summarized in a report.

The survey should only be completed only once.

If possible, please answer the questions in English. Answers written in German will also be accepted.

DEMOGRAPHIC INFORMATION

First, we would like to ask you a few questions about you.

1. You are: male/female

2. You are:
   - 18- 35 years
   - 36-50 years
   - 51-65 years
   - 65+ years
   - age not specified

3. What is the highest qualification that you have?

   - High school

- Bachelor
- Master
- Magister
- German Diploma
- State examination
- Doctorate
- Professor
- Habilitation
- Other:

4. Your current position is:

- Professor
- Researcher
- Lecturer
- Principal Investigator / Head of laboratory
- Postdoc
- Student
- Technician / Staff
- Administration
- Management
- Corporate Social Responsibility/Public Relations
- Other (please specify):

5. Your employer/institution is:

- Public
- Private
- Public/private partnership
- Not applicable/specified
- Other (please specify)

6. Which best describes your employer/institution or your work sector? (multiple answers possible)

- Academic: non-commercial research
- Academic: teaching focused
- Government
- Collection
- Commercial research
- Industry (production)
- Industry association

- Non-profit organization
- Other (please specify):

7. From which perspective are you completing this survey:

   - Your individual perspective
   - On behalf of your employer/institution
   - On behalf of an association and/or its members
   - Unsure/Not applicable

8. If you are involved in research, which of the following best describes the funding agency for your research? (multiple answers possible)

   - European Union
   - Federal/national government
   - State/regional government
   - Local government
   - Corporation
   - Private foundation
   - Other (please specify):

9. For researchers: Which best describes your research discipline? (multiple answers possible)

   - Agronomy (food, feed, fuel etc.)
   - Animal breeding
   - Animal health
   - Biochemistry
   - Biocontrol (herbicides, insecticides, fungicides etc.)
   - Bioinformatics
   - Biology
   - Biostimulants
   - Biotechnology
   - Chemistry
   - Computer science
   - Ecology
   - Engineering
   - Evolutionary biology
   - Forestry
   - Genetics

- Genetic engineering
- Genomics
- Health sciences
- Information science/technology
- Medicine
- Metabolomics
- Metagenomics
- Microbiology
- Pharmacy
- Phenomics
- Plant breeding
- Plant pathology
- Proteomics
- Soil science
- Taxonomy
- Transcriptomics
- Other (please specify):

10. Only for commercial/industry: Which sector best describes your work sector? (multiple answers possible)

- Animal breeding
- Plant breeding
- Biotechnology
- Biocontrol
- Biostimulants
- Cosmetics
- Food
- Feed
- Pharmaceuticals
- Other (please specify):

DIGITAL SEQUENCE DATA OR DIGITAL SEQUENCE INFORMATION?

The next group of questions refers to your perception of the difference between digital sequence data and digital sequence information.

11. Do you think there is a difference between digital sequence data and digital sequence information?

- Yes
- No

12. In your opinion, what is the difference?

13. How would it be possible to determine the difference between the two, i.e. how would you define the cut-off point?

14. How strongly do you agree or disagree with the following statements? (agree strongly, agree somewhat, neither agree nor disagree, disagree somewhat, disagree strongly, not sure)

    - Digital sequence data are only a collection of letters
    - Digital sequence data can convey meaning
    - Digital sequence data must be interpreted before they have meaning
    - Digital sequence data must be interpreted before they have value
    - Interpretation of digital sequence data creates digital sequence information
    - Only experts can create digital sequence information
    - Digital sequence information is not more valuable than digital sequence data

<u>USE OF DIGITAL SEQUENCE INFORMATION/DATA</u>

The next set of questions refers to the use of digital sequence information or digital sequence data. Here, the terms data and information are used interchangeably.

15. Do you use digital sequence information/data from genetic resources in your research/work?
    - Yes
    - No

16. Are you likely to use digital sequence information/data from genetic resources in your research/work in the future?
    - Yes
    - No

17. Which types of digital sequence information/data from genetic resources do you typically use for your work/research? (multiple answers possible)

    - Animal
    - Plant
    - Fungi
    - Algae
    - Archaea
    - Bacteria
    - Virus

- Other

18. Do you know from which country the original genetic resources underlying the sequences come from?

- Always
- Often
- Usually
- Sometimes
- Rarely
- Never
- N/A

19. Where do the original genetic resources underlying the sequences come from? (multiple answers possible)

- Germany
- Other European Union member states
- Other European countries (not European Union)
- Africa
- Asia
- Australasia and Pacific
- North America
- South America
- Antarctica
- Arctic
- Ocean
- Other (Please specify)
- Not applicable

20. If you don't know where the original genetic resources underlying the sequences come from, why not? (multiple answers possible)

- There are no metadata about the origin of the genetic resources
- There are metadata about the origin of the genetic resources but I have no access to them
- There are metadata about the origin of the genetic resources but they are not relevant to my work/research
- I use too much genetic sequence information/data to keep track of where they all the genetic resources originally come from
- Don't know/unsure
- Other (please specify):

21. Is it possible for you to generate synthetic genetic resources from the digital sequence information/data?
    - Yes
    - No

22. Is it possible for another laboratory to generate synthetic genetic resources from digital sequence information/data on your behalf?
    - Yes
    - No

23. Could you, at least in theory (or in your opinion in the future), work/research only using synthetic genetic resources produced from digital sequence information/data?
    - Yes
    - No

DIGITAL SEQUENCE INFORMATION/DATA GENERATED IN THE COURSE OF YOUR WORK/RESEARCH

24. Are digital sequence information/data generated from genetic resources through your work/research? (e.g. either you generate the digital sequence information/data yourself, it is generated on your behalf by another laboratory, someone in your research consortium generates it)
    - Yes
    - No

25. Approximately how much digital sequence information/data are generated each year in the course of your work/research?

    - Less than 50 sequences
    - 50 to 100 sequences
    - Hundreds of sequences
    - Thousands of sequences
    - Millions of sequences
    - Billions of sequences

26. The amount of digital sequence information/data generated in the course of your work research:

    - is staying the same
    - is decreasing
    - has been increasing slowly since 2012
    - has been increasing steadily since 2012
    - has been increasing rapidly since 2012

27. How important are digital sequence data/information from these genetic resources to your work/research?

- Very important
- Somewhat important
- Neither important nor unimportant
- Somewhat unimportant
- Not important at all
- Not sure

28. How do you use the digital sequence information/data generated by you (or on your behalf) for your work/research? Please provide a brief description.

29. Would it be possible to do your work/research without generating digital sequence information/data (or having it generated on your behalf)?
    - Yes
    - No

30. Why/why not?

31. Once the digital sequence information/data has been generated (either by you or on your behalf by another laboratory etc.), who usually has access to it? (multiple answers possible)

    - Only me
    - My research team
    - My institute/company
    - My collaborators at other institutes
    - The sequencing laboratory
    - Other (please specify):

32. Have you ever shared your genetic sequence information/data with anyone outside your research group, collaborators or institute?

    - Yes
    - No
    - Not sure
    - Not applicable

33. If yes, how did you share this digital sequence information/data? (multiple answers possible)

    - Email / direct contact
    - Website (personal or laboratory/institute's website)
    - Through a network
    - Don't know
    - Other (please specify):

34. Approximately how much of your digital sequence information/data has been shared with others outside your research team/institute/collaborators?

    - Less than 25%
    - 25-50%
    - 50-75%
    - 75-100%

35. Have you ever published your digital sequence information/data? Yes/No

36. If so, where did you publish this digital sequence information/data? (multiple answers possible)

    - Publicly accessible database
    - Database with only registered/limited access
    - Journal repository as supplementary material
    - Discipline-based repository
    - Other (please specify):

37. Please list the databases, repositories etc. where you usually publish your digital sequence information/data.

38. Approximately how much of your digital sequence information/data has been published?

    - Less than 25%
    - 25-50%
    - 50-100%
    - 75-100%

39. When do you typically share or publish your digital sequence information/data?

    - Immediately after sequencing
    - Before I publish the results of my research
    - After I publish my results of my research
    - Not applicable
    - Other (please specify):

40. How important is it for you to be able to share your digital sequence information/data with others? (very important, somewhat important, neither important nor unimportant, somewhat unimportant, not important at all, not sure)

41. How important is it for you to be able to publish your digital sequence information/data?

    - Very important

- Somewhat important
- Neither important nor unimportant
- Somewhat unimportant
- Not important at all
- Not sure

42. How important are the following motivators for you in terms of sharing/publishing the genetic sequence information/data generated through your research/work? (very important, somewhat important, neither important nor unimportant, somewhat unimportant, not important at all, not sure)

    - Requests from other users of digital sequence information/data
    - It is the policy of your institution to share research results
    - Publishers require you to share/publish my research results
    - Funders require you to share/publish my research results
    - It is part of the practice/policy of the scientific community to share research results
    - Dissemination and recognition of your work/research
    - Acceleration of scientific research
    - Acceleration of the possible application of scientific research
    - Personal commitment to open data

42a. Other (please specify):

43. Do you ever have concerns about liability or rights of making genetic sequence information/data available to people outside your institution/research group/collaborators?

    - Yes
    - No

44. Why/why not?

45. Would it be possible for you to know who has accessed your digital sequence information/data after it was published?

    - Yes
    - No

46. Why/why not? Please be specific, e.g. it is not possible to trace it.

47. Would it be possible to know how your published digital sequence information/data is used after it is accessed?
    - Yes
    - No

48. Why/why not?

49. To your knowledge, has your shared/published digital sequence information/data ever been reused by anyone outside your institution/research group?

    - Yes
    - No
    - Don't know
    - Not applicable

50. If yes, how was your shared/published digital sequence information/data reused? (multiple answers possible)

    - It was reanalysed
    - It was incorporated into a larger data set
    - It was used to draw new conclusions
    - It was used to develop a product
    - Other (please specify):

You are over halfway through the survey!

<u>USING DIGITAL SEQUENCE INFORMATION/DATA FROM OTHER SOURCES</u>

Next, we would like to ask you about how you use digital sequence information/data from other sources, i.e. digital sequence information/data that were not generated by you or on your behalf.

51. Do you use digital sequence information/data generated that have been generated by people outside your research group/collaborators/institute?
    - Yes
    - No

52. How do you use the digital sequence information/data generated by others for your work/research? Please provide a brief description.

53. Tell us how much you agree or disagree with the following statements (agree strongly, agree somewhat neither agree nor disagree, disagree somewhat, disagree strongly, not sure)

    - Lack of access to digital sequence information/data would be a major impediment to my research/work
    - Lack of access to digital sequence information/data would be would restrict my ability to do my work/research.
    - Lack of access to digital sequence information/data would be would not be a problem for me.

54. Roughly how much digital sequence information/data (not generated through your own work/research) would you use each year?

- None
- Less than 100 sequences
- Hundreds of sequences
- Thousands of sequences
- Millions of sequences
- Billions of sequences

55. The amount of digital sequence information/data you use which are generated by others:

- is staying the same
- is decreasing
- has been increasing slowly since 2012
- has been increasing steadily since 2012
- has been increasing rapidly since 2012

56. How do you get the digital sequence information/data which you do not generate yourself through your work/research? (multiple answers possible)

- Email / direct contact
- Personal or laboratory/institute's website
- Publicly accessible database
- Database with only registered/limited access
- Journal repository as supplementary material
- Discipline-based repository
- Other (please specify):

57. List the databases/ repositories etc. from which you typically get published genetic sequence data/information.

58. Would it be possible for you to do your research only using digital sequence information/data you get from other sources, i.e. there would be no need for the underlying genetic resources?
    - Yes
    - No

59. Why/why not? Please be specific, e.g. the nature of your work, regulatory requirements etc.

60. When you download genetic sequence information/data from a database/repository, do you know where the original genetic resource underlying the sequence comes from?
    - Yes
    - No

61. Why/Why not?

62. Do you know of any conditions/user agreements associated with using digital sequence information/data you get from other sources? Yes/No

63. If yes, what types of conditions are there on the use of digital sequence information/data you get from databases/repositories?

    - The digital sequence information/data cannot be shared with others
    - The digital sequence information/data cannot be published without permission of another person
    - The digital sequence information/data cannot be used for commercial purposes
    - The digital sequence information/data may be used for commercial purposes but a license fee must be paid
    - Don't know
    - Other (please specify):

64. Researchers: Have you ever republished the digital sequence information/data that you have obtained from other sources?
    - Yes
    - No
    - Not sure
    - Not applicable

65. Industry: Have you ever used digital sequence information/data obtained from other sources to create a product?

    - Yes
    - No
    - Not sure/Not applicable

66. Which attributes do you think are most important for digital sequence information/data? (agree strongly, agree somewhat neither agree nor disagree, disagree somewhat, disagree strongly, not sure)

    - It is easy to access to digital sequence information/data
    - Access to digital sequence information/data is unrestricted
    - I can re-use the digital sequence information/data without restrictions
    - I can republish the digital sequence information/data
    - Any license terms relating to use of the digital sequence information/data are clear
    - Access to digital sequence information/data is low cost

- Access to digital sequence information/data is free of charge

ACCESS (PRIOR INFORMED CONSENT AND MUTUALLY AGREED TERMS)

67. Are you familiar with the term “access and benefit sharing” under the Nagoya Protocol?

    - Yes
    - No

68. Have you ever obtained Prior Informed Consent (PIC) and entered into Mutually Agreed Terms (MAT) in order to access and utilize genetic resources from another country?

    - Never
    - Seldom
    - Sometimes
    - Often
    - Always
    - Not applicable

69. Have your PIC and/or the MAT ever made any reference to sequencing of the genetic resource/s?
    - Yes
    - No

70. If yes, was sequencing of the genetic resources permitted?

    - Never
    - Seldom
    - Sometimes
    - Often
    - Always
    - Not applicable

71. If sequencing was not permitted, why not:

72. If you were not permitted to sequence genetic resources, what were the consequences for your work/research?

73. Were you allowed to share or publish the digital sequence information/data you generated?

    - Never
    - Seldom
    - Sometimes
    - Often
    - Always

- Not applicable

74. If you were not permitted to publish the digital sequence information/data, why not?

75. If you were not permitted to publish the digital sequence information/data, what were the consequences for your work/research?

76. What would be the consequences for your work/research if, in future, PIC and MAT will not allow you to sequence the genetic resources or to share/publish the digital sequence information/data?

ACCESS AND BENEFIT SHARING

77. How much do you agree or disagree with the following statements? (agree strongly, agree somewhat neither agree nor disagree, disagree somewhat, disagree strongly, not sure)

- It is fair to give something in exchange for having access to digital sequence information/data
- Countries that provide genetic resources should get something when digital sequence information/data from those genetic resources are used for basic research
- Countries that provide genetic resources should get something when digital sequence information/data from those genetic resources are used for commercial purposes
- I would pay in exchange for being able to access and use digital sequence information/data
- I would pay in exchange for being able to access and use digital sequence information/data only if a commercial product is developed from it
- It would be possible to pay a fee every time I access digital sequence data/information
- It would be possible to pay a fee for being able to access and use as much digital sequence data/information as I want within a specific timeframe e.g.one year
- I think it is fair to give research funding to countries that provide original genetic resources in exchange for access to digital sequence information/data
- I think it is fair to support scientific research and development in countries that provide genetic resources in exchange for access to digital sequence information/data
- It would be fair to transfer knowledge and technology to countries that provide genetic resources in exchange for access to digital sequence information/data

CONCLUSION

78. Please feel free to share any additional comments, opinions, questions, or suggestions you may have about digital sequence information/data e.g. its use and importance to your work and research, potential access and benefit sharing arrangements etc.

The questionnaire is now complete.

Thank you very much for your time!