

# Open access: a technical assessment for the debate on benefit-sharing and digital sequence information

Rodrigo Sara<sup>1,2</sup>, Andrew L. Hufton<sup>1</sup>, Scarlett Sett<sup>1</sup>, Amber H. Scholz<sup>1\*</sup>

<sup>1</sup> Leibniz Institute DSMZ, Braunschweig, <sup>2</sup> One Planet Solutions  
\*correspondence to [amber.h.scholz@dsmz.de](mailto:amber.h.scholz@dsmz.de)

FINAL VERSION December 22, 2021

**Summary:** *The scientific community has a 40-year history of open access to sequence data. This study examines how, when, and why open access arose and the working principles of open access publication and open data. It further explores the international policy landscape on open access and concludes with an assessment of the bi-directional policy implications between open access and the debate around digital sequence information.*

This report is a deliverable for the project "Mögliche Optionen für die Behandlung von DSI im Spannungsfeld der 3 CBD-Ziele" with the Grant Number 3519532011, led by the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), which has been commissioned by the German Federal Agency for Nature Conservation with funds from the German Federal Ministry for the Environment, Nature Conservation, Nuclear Safety and Consumer Protection. The responsibility for the content of this publication lies with the authors.

## Table of Contents

Executive summary.....	3
Introduction.....	5
“Open access” and the rationale of “openness” .....	6
The open access publishing ecosystem.....	8
Open data in scholarly research .....	11
Open DSI and ABS .....	13
DSI and public databases .....	13
DSI and ABS policy tensions .....	14
What can the INSDC’s access policy tell us about open access for DSI?.....	14
Can controlled and/or differentiated access provide an alternative to open access to DSI?....	17
What do international, regional, and national policies tell us about open access for DSI? .....	19
UNESCO.....	20
European Union .....	20
African Union.....	22
Brazil .....	22
India.....	23
China .....	23
Conclusion .....	26
References.....	28

---

## Executive summary

---

Access and benefit benefit-sharing (ABS) associated with digital sequence information (DSI) on genetic resources remains a contentious issue in the lead-up to the 15th meeting of the Parties to the UN Convention on Biological Diversity. A compromise on this issue will likely be necessary for the Parties to reach consensus and adopt the Global Biodiversity Framework. Sequence data are overwhelmingly available under open, unrestricted conditions and the broader DSI “data ecosystem” inherently depends on this openness. This policy report examines what open access to DSI means and intends to add technical and historical perspectives to the DSI policy debate.

We evaluate open access first in the context of the broader open science movement, which represents an emerging, cooperative approach to the scientific process based on new ways of diffusing knowledge in a frictionless manner. In this context, the reuse of data and knowledge is strongly promoted using digital technologies and new collaborative tools. We complement this with an overview of key elements in the open “ecosystem”: open publishing models, standard open licenses, definitions of “open” and “open access”. These elements facilitate and enable access and use of the outputs of scientific research. We discuss in particular detail the International Nucleotide Sequence Database Collaboration (INSDC) and its policy of “free and unrestricted access” to show that public databases that provide access to DSI are permissive and unrestricted at present.

Interoperability is also a defining characteristic of access and reuse of DSI across a vast public biological database ecosystem. DSI interoperability has both technical and legal dimensions that underpin value-creation in the life sciences sector. The life sciences span fields ranging from biodiversity, conservation, evolution, ecology, agriculture, global and public health research through to the commercial areas of pharmaceuticals, cosmetics, nutrition, breeding, food and industrial processing. Across these fields, researchers must routinely combine, integrate and compare large sets of sequences and related information in order to gain insights that drive forward basic research and innovation. The ubiquitous availability of DSI, which fuels life sciences research and innovation, is enabled by near uniform terms of access and permissive conditions on reuse across databases. This underlying legal interoperability, in effect, maintains a vast DSI data commons as a global public good by minimizing transaction and other costs the scientific community would otherwise experience in accessing and reusing these data.

More broadly, international science is clearly moving towards greater openness and data sharing, and national and regional governments are increasingly promoting open data and open science to inform science and innovation policy. At the same time, a more nuanced approach to “open” is emerging, which calls for access to research results to be “as open as possible” in recognition that legitimate circumstances may call for “closed” or “regulated” access, for example, concerning national security, confidentiality, intellectual property (IP) rights, privacy or research ethics. In November 2021, UNESCO parties voted to affirm a Recommendation on Open Science in which timely, free and open access to research data, subject to access restrictions where proportionate and justified, is considered a foundational pillar. The Recommendation seeks to build a coherent and internationally aligned vision of open science to support scientific cooperation and to make science more transparent, accessible, equitable and inclusive. This overall trend towards open is not just a catalyst to research, innovation and technological development, but also an integral feature of sustainable economic development governance.

We also consider the Global Biodiversity Information Facility (GBIF)), which facilitates “open differentiated access” to biodiversity data using standard open licenses. Other well-known “controlled-access” examples, including the Global Initiative on Sharing Avian Influenza Data (GISAID), which provides access to a limited pool of viral genomic data, are assessed. Such approaches create substantial interoperability obstacles that would be likely to disrupt research and innovation if applied on a global scale. We also note the potential for some systems, for example, licenses, to undermine benefit-sharing objectives because they readily facilitate avoidance behavior and jurisdiction shopping of least-restrictive terms.

We conclude that open access and benefit-sharing can be mutually reinforcing rather than in conflict and make three key observations in this regard. First, recent efforts to define or characterize “open” in the context of scientific research data, for example, the Panton Principles for Open Data in Science, the FAIR Data Principles and the UNESCO Open Science Recommendation, should serve as a starting point for characterizing or defining open access in the context of DSI. Second, whether a policy option maintains open access to DSI in public databases should be evaluated based on the extent to which it is open (everyone can see the data), and technically and legally interoperable (machines and people can transfer and electronically move and manipulate the data). These two primary characteristics are the engine that drives the generation of non-monetary and monetary value from DSI. Third, a multilateral and universal mechanism that deals with the global DSI dataset as a whole – such as global terms and conditions if benefit-sharing is coupled to the use of DSI, micro-levy or other payments decoupled from the use of DSI – seems best suited to minimize frictions to data flow and interoperability that could hinder scientific research and innovation.

## Introduction

---

The 14<sup>th</sup> Conference of the Parties to the UN Convention on Biological Diversity (CBD COP14) acknowledged a divergence of views regarding benefit-sharing from the use of digital sequence information (DSI) on genetic resources.<sup>1</sup> Access and benefit-sharing (ABS) policy associated with DSI has remained a contentious issue in the lead-up to COP 15<sup>2</sup>, which is expected to adopt a Global Biodiversity Framework (GBF) to serve as a 10-year strategy to implement the 2050 biodiversity vision of living in harmony with nature. A decision on DSI is anticipated at COP 15, and a compromise on this issue will likely be necessary for the Parties to reach the consensus needed to adopt the GBF.

Open access to biological data in public databases is at the heart of the DSI debate, both in terms of its perceived role in either undermining benefit-sharing or being itself a form of benefit-sharing. More recently, evolving DSI policy discussions under the CBD are exploring the development of an ABS framework that preserves open access to DSI through public databases while ensuring benefit-sharing arising from its use.<sup>3</sup> Some have suggested that such a framework should not prevent or unduly restrict free and open access to digital sequence information in public databases nor significantly hinder scientific research and innovation.<sup>4</sup>

Despite its growing prominence in the DSI debate, there has been limited discussion to date as to what constitutes “open access” to biological data in public databases and whether the term is even capable of a precise definition. This challenge was recently acknowledged at the 3<sup>rd</sup> meeting of the Open-Ended Working Group (OEWG) of the post-2020 Global Biodiversity Framework in which the Co-Leads of a Contact Group on DSI observed that while there appears to be convergence on the importance of open access to DSI in general, interpretations of “open access” differ.<sup>5</sup>

It is imperative that DSI policy actors are able to compare “apples to apples” when discussing the role and nature of open access in the ongoing deliberations, which are approaching a critical phase. A common understanding regarding the criteria, constituent elements, or boundaries of open access is a necessary precursor for evaluating whether ABS-related restrictions or obligations are consistent with open access to DSI. It will also help evaluate what kinds of changes to the status quo concerning access to biological data in public databases have the potential to hinder scientific

---

<sup>1</sup> Conference of the Parties to the Convention on Biological Diversity, “14/20. Digital Sequence Information on Genetic Resources.”

<sup>2</sup> Initially scheduled to take place late-2020 and delayed due to the ongoing COVID-19 pandemic. COP 15 was opened in a virtual format in October 2021 with substantive meetings to take place in 2022.

<sup>3</sup> Open access featured prominently in the informal discussion forum and related webinars focused on detailing and evaluating DSI policy options, which were organized by the Secretariat of the CBD under the leadership of the Co-Chairs of the Open-Ended Working Group (OEWG) of the post-2020 Global Biodiversity Framework (see <https://www.cbd.int/dsi-gr/forum.shtml>). It also featured prominently in the deliberations of a contact group established to consider DSI at the third meeting of the OEWG3 on 23 August – 2 September 2021 (see <https://www.cbd.int/article/new-dates-oweg3-august-september-2021>).

<sup>4</sup> See, for example, the Background Paper on DSI prepared by the CBD Secretariat (Open-ended Working Group on the Post-2020 Global Biodiversity Framework, “Digital Sequence Information on Genetic Resources.”)

<sup>5</sup> It also noted that “divergent views were expressed regarding ‘open access’ to DSI and whether access should be free, restricted or unrestricted, regulated or unregulated, subject to free, prior and informed consent for at least some groups, such as indigenous peoples and local communities, for all, or not permitted at all.” See Tshitwamulomoni and Voigt-Hanssen, “Co-Leads’ Summary of the Discussion of the Contact Group Regarding Areas of Potential Convergence and of Apparent Divergence on Digital Sequence Information on Genetic Resources.”

research and innovation that relies on DSI.<sup>6</sup>

This policy brief provides an analysis of “open access” to inform ongoing DSI policy discussions in the final stretch to COP15.<sup>7</sup> It is designed to equip actors in the DSI policy debate to evaluate and discuss open access in a more nuanced manner and to support further inquiry on how open access to DSI can facilitate benefit-sharing without significantly hindering scientific research and innovation. We begin the brief at the meta-level of openness and work our way through to DSI.

## “Open access” and the rationale of “openness”

The term “open access” was articulated for the first time by the 2002 Budapest Open Access Initiative, referring specifically to online access to research literature that was free of charge and free of most usage restrictions.<sup>8</sup> In the following two decades, the terms “open” and “open access” have frequently been used interchangeably and have been applied in a broad range of contexts associated with scientific research beyond academic publishing. Given these developments, it is useful to consider the context in which the term arose, its evolving use and the underlying rationale of “openness”.



Figure 1: Benefits of open access. Reused from Bauer et al., “Recommendations for the Transition to Open Access in Austria.” Distributed under [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/).

The emerging “open” movement in scientific research builds on principles of “open access” established in relation to online publishing of scientific journals and “open source” established in

<sup>6</sup> Such an evaluation is beyond the scope of this report, however, it is expected to benefit from this report and to be the logical next-step of inquiry by the authors.

<sup>7</sup> Decision 14/20 by the Conference of the Parties at COP 14 (noted in [1] above) implemented a science- and policy-based process on digital sequence information to resolve the divergence of views among Parties regarding benefit-sharing from the use of digital sequence information on genetic resources.

<sup>8</sup> <https://www.budapestopenaccessinitiative.org/read>; see also Suber, “Gratis and Libre Open Access.”

relation to software licensing, both of which gained prominence in the 1990s as access to the internet became widely available and online dissemination became the norm.<sup>9</sup> The notion of “open” encompasses a wide range of practices seeking to ensure free, online access to information, including the outputs of scientific research.

While these movements have different histories and practices, they often share a recognizably common conception of what “open” means. For example, the Open Knowledge Framework developed a definition of “open”, based on community feedback and consultation, that can be applied to different content types spanning the various open movements. Content is considered open “if anyone is free to access, use, modify, and share it – subject, at most, to measures that preserve provenance and openness”,<sup>10</sup> as noted in further detail below.

While the open movements focus on sharing content with relatively few restrictions, this does not mean that downstream users are free of all obligations. In particular, open movements emphasize preserving “provenance” to ensure attribution and credit for the original content creators or copyright owners.

Within the scholarly domain, this broader open movement has coalesced into “open science”, a still evolving concept that embodies “openness” in a wide range of contexts spanning research publications, open research data, open source software, open collaboration, open peer review, open notebooks, and open educational resources.<sup>11</sup> Collectively it represents an emerging approach to the scientific process based on cooperative work and new ways of diffusing knowledge frictionlessly and promoting its reuse using digital technologies and new collaborative tools.

Advocates of open science argue more broadly that scholarly knowledge is a public good, rendered useful only when it is shared and utilized.<sup>12</sup> They point to open practices as a research accelerator that fuels scientific discoveries and innovation,<sup>13</sup> and far-reaching benefits to society at large (Fig. 1). In the context of DSI, a combination of good scientific practice, growing societal pressure for transparency and ethics in scientific discovery, and open-access requirements by funding agencies led to the now near-universal scientific practice of submitting newly generated nucleotide sequence data to the INSDC.<sup>14</sup>

The economic arguments for open access are also persuasive. A 2014 report “Open for Business” commissioned by the Omidyar Network<sup>15</sup> evaluated the business case for open data and estimated that economic growth and productivity from open data would add an additional \$3.2 trillion in global GDP with governments, the private sector, individuals and communities standing to benefit. The innovation and information from open data will inform investment, drive the creation of new

---

<sup>9</sup> [https://cshl.libguides.com/open\\_access/history\\_policy](https://cshl.libguides.com/open_access/history_policy). Adapted from Tennant et al., “The Academic, Economic and Societal Impacts of Open Access.”

<sup>10</sup> Available at <https://opendefinition.org/od/2.1/en/>

<sup>11</sup> See this UNESCO brochure for a useful overview [https://en.unesco.org/sites/default/files/open\\_science\\_brochure\\_en.pdf](https://en.unesco.org/sites/default/files/open_science_brochure_en.pdf)

<sup>12</sup> Bauer et al., “Recommendations for the Transition to Open Access in Austria.”

<sup>13</sup> Woelfle, Olliaro, and Todd, “Open Science Is a Research Accelerator.”

<sup>14</sup> Rohden et al., “Combined Study on Digital Sequence Information in Public and Private Databases and Traceability.”

<sup>15</sup> SPARC quotes the report at <https://sparcopen.org/open-data/>. The report is no longer available at the source referenced by SPARC ([https://omidyar.com/sites/default/files/file\\_archive/insights/ON%20Report\\_061114\\_FNL.pdf](https://omidyar.com/sites/default/files/file_archive/insights/ON%20Report_061114_FNL.pdf)) however, a summary of the report by the Open Knowledge Foundation is available at <https://blog.okfn.org/2014/06/23/the-business-case-for-open-data/>.

industries, and inform decision making and research. Similarly, a 2020 report evaluating “Economic Impact of Open Data” in the European Union<sup>16</sup> estimated the value of the open data market in Europe as €184.45 billion in 2019. This is forecast to grow to up to €334 billion by 2025 with significant growth in a range of sectors driven by increased innovation in science and technology, efficiency gains and savings costs resulting from open data. While economic data or modelling focusing specifically on open access to biological databases, let alone just DSI, is hard to come by,<sup>17</sup> the ubiquitous availability of genetic sequence and related data have accelerated the growth of life sciences research and innovation. The synthetic biology market, for example, which relies heavily on open DSI in early phases of development, is projected to grow from USD \$9.5 billion in 2021 to \$30.7 billion by 2026.<sup>18</sup>

## The open access publishing ecosystem

---

To gain a deeper understanding of open access, the publishing of scientific publications can be quite instructive. Launched in 1993, the arXiv repository was developed by physicists to freely share “preprint” versions of scientific papers. This repository was seen as a trailblazer in open access publishing<sup>19</sup> and demonstrated that publications could be shared cheaply and freely around the globe via web distribution, whetting the scholarly community’s appetite for rapid, barrier-free digital access.

In 2000, an influential group of biologists, including a Nobel Laureate – Harold Varmus, Patrick Brown and Michael Eisen – became convinced that this new trend represented the future of scholarly publication. They published an open letter declaring they “will publish in, edit or review for, and personally subscribe to only those scholarly and scientific journals that have agreed to grant unrestricted free distribution rights to any and all original research reports that they have published.”<sup>20</sup> Journal options specifically dedicated to open access publishing began to emerge.

The term “open access” was coined by the 2002 Budapest Open Access Initiative, referring to open access as: “online access to research literature that was free of charge and free of most usage restrictions.”<sup>21</sup> Core components of open access publishing were further refined by the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities<sup>22</sup> in 2003 and the Bethesda Statement on Open Access Publishing in 2003.<sup>23</sup> These three documents – collectively known as the BBB declarations<sup>24</sup> – define open access to mean the “*free, irrevocable, worldwide, perpetual right of access to, and a license to copy, use, distribute, transmit and display the work publicly and to make and distribute derivative works, in any digital medium for any responsible purpose, subject to proper attribution of authorship.*” These milestones in open access publishing

---

<sup>16</sup> Publications Office of the European Union., *The Economic Impact of Open Data*.

<sup>17</sup> A notable U.S. study published in 2013 estimated that by 2012 the research, development and commercial activities that leverage the investment in the human genome sequencing projects directly and indirectly generated nearly USD \$1 trillion in total economic impact since 1988 including USD \$65 billion in US economic output and USD \$31 billion towards U.S. GDP in 2012 alone; see <http://www.unitedformedicalresearch.com/wp-content/uploads/2013/06/The-Impact-of-Genomics-on-the-US-Economy.pdf>.

<sup>18</sup> According to MarketsandMarkets, “Synthetic Biology Market - Global Forecast to 2026.”

<sup>19</sup> Moore, “A Genealogy of Open Access.”

<sup>20</sup> <https://plos.org/open-letter/>

<sup>21</sup> <https://www.budapestopenaccessinitiative.org/read>; Suber, “Gratis and Libre Open Access.”

<sup>22</sup> <https://openaccess.mpg.de/Berlin-Declaration>

<sup>23</sup> <http://legacy.earlham.edu/~peters/fos/bethesda.htm>

<sup>24</sup> <https://crai.ub.edu/en/crai-services/intellectual-property/open-access-ub/what-is>

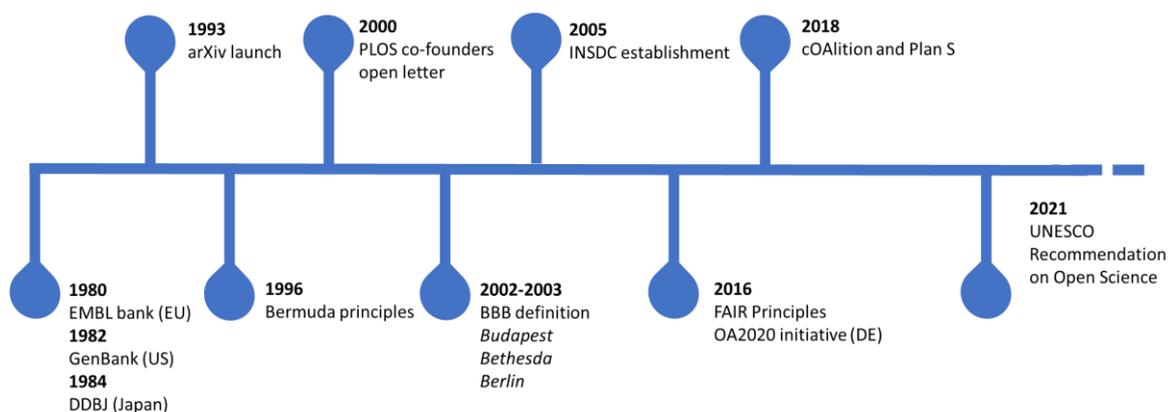


Figure 2: Overview of key milestones in the open movement

are depicted in Fig. 2, which also incorporates other milestones mentioned later in this brief that relate to the INSDC and the open movement more broadly.

Since the term open access was coined, thousands of journals have adopted policies that embrace some or all of the core components of open access, and specialized terms have evolved to describe different publishing modalities for achieving open access, such as through self-archiving (i.e. “green”) and open-access journals supporting by author fees (i.e. “gold”).<sup>25</sup> Borrowing the terms ‘gratis’ and ‘libre’ from the open source software domain has been used as a convenient way to distinguish between open access modalities that provide online access free of charge but subject to all or most copyright restrictions (gratis) compared with those that provide online access free of charge and free of most restrictions (i.e. libre). Gratis open access is considered to insufficiently embrace the core components of open access under the BBB Declarations, whereas libre open access captures a broad spectrum of access modalities.<sup>26</sup>

To help navigate this fragmentation, SPARC published an open access guide that standardizes open access terminology and consolidates the key elements of open access publishing of journal policies. They outline six fundamental aspects of open access and how these aspects are implemented across the spectrum between open and closed access:<sup>27</sup>

1. Reader rights,
2. Reuse rights,
3. Copyrights,
4. Author posting rights,
5. Automatic posting,
6. Machine readability.

This is intended to move the conversation beyond the deceptively simple question of “Is It Open Access?” toward an evaluation of “How Open Is It?” A similar line of inquiry may prove useful to

<sup>25</sup> In fact, open access publishing modalities are further fragmented, and many more colour variant names have been invented. The DePaul University Library provides a convenient breakdown of so called, Green, Blue, Yellow, Gold, White, Black and also hybrid models of Open Access: <https://libguides.depaul.edu/c.php?g=844896&p=6039089>

<sup>26</sup> Suber, “Gratis and Libre Open Access.”

<sup>27</sup> Ibid.



Figure 3: Breakdown of Creative Commons Licenses. Reused from foter, “[How To Attribute Creative Commons Photos – Foter Blog](#).” Distributed under [CC BY-SA 3.0](#)

characterize open access in a DSI context, for example, regarding provenance, reuse rights and machine readability and interoperability.<sup>28</sup>

While the open access publishing landscape is diverse, there have been areas of remarkable standardization. Most open access publications are released under one of the six licenses offered by Creative Commons. These licenses take inspiration from open-source (i.e. computer source code) licensing, which permits software to be freely used, modified, and shared according to a set of criteria:<sup>29</sup> free distribution of source code; permits derived works if distributed on the same terms; maintains the integrity of authors’ source code; imposes no discrimination against groups or persons, or fields of endeavor; allows distribution without an additional license; license must not be specific to a product or restrict other software; and must be technology neutral. Despite these generally accepted criteria, by some estimates, over 200 open source licenses now exist<sup>30</sup> with vastly different levels of adoption and interoperability, thus creating a splintered software commons that places legal risk and transaction costs on users seeking to incorporate software code sourced

<sup>28</sup> A draft book chapter provides a preliminary analysis along these lines. A pre-print is available; see Sara, Hufton, and Scholz, “Compatible or Incompatible? DSI, Open Access, and Benefit-Sharing.”

<sup>29</sup> As per the Open Source Definition developed by the Open Source Initiative (OSI) (see <https://opensource.org/osd-annotated>), which maintains a list of approved licenses that comply with the criteria, consisting of a little over 80 open source licenses (see <https://opensource.org/licenses>).

<sup>30</sup> <https://www.whitesourcesoftware.com/resources/blog/open-source-licenses-explained/>

under different open licenses.<sup>31</sup>

The Creative Commons suite of licenses provide an orderly breakdown of the rights of reuse under libre open access, subject at least to the right of attribution and at most to restrictions on one or more of commercial use, making derivatives, and the obligation to share a copy or derivative on identical terms (Fig. 3). An additional option, which does not operate as a license, CC0, releases a work into the public domain without legal attribution obligations (although attribution in the scholarly domain is still maintained by community ethical norms). First released in 2002, over one billion works were available under Creative Commons by 2015 and are estimated to now exceed two billion.<sup>32</sup>

The Open Knowledge Foundation maintains a list of conformant and non-conformant licenses. The latter includes, for example, Creative Commons licenses with non-commercial (NC) or no derivatives (ND) designations that are deemed inconsistent with the Open Definition. OASPA strongly promotes the use of the CC BY variant and disallows the use of the some of the more restricted versions among its member organizations.<sup>33</sup> This is also acknowledged by Creative Commons, which notes that some of its licenses may be inconsistent with open access under international initiatives.<sup>34</sup> This distinction may, however, be lost on the broader scientific research community given that Creative Commons licenses (including those that restrict commercial use or derivatives) are frequently touted as promoting, facilitating or empowering open access and open science in a generalized manner.<sup>35</sup>

Overall, while there is substantial diversity in open access publishing, there are commonly accepted definitions of “open access”, grounded in the original BBB declarations, and centralizing forces that have succeeded in standardizing licensing to a large degree across publishers. Nonetheless, as in other contexts in which “open” is employed, a spectrum is evident with fuzzy boundaries or criteria for less open options that nevertheless qualify as “open”.

## Open data in scholarly research

According to SPARC, open access to data is an important complement to open access publishing. It considers research to include “open data” if the data are freely available on the internet, permits any user to download, copy, analyzes, re-process, pass to software or use for any other purpose, and is without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself.”<sup>36</sup> Its further guidance indicates that a key element to open data is legal certainty to redistribute and fully reuse data. They suggest this should be achieved by a license that

---

<sup>31</sup> See for example the following, which include diagrams depicting open source license compatibility issues:

<https://github.com/HansHammel/license-compatibility-checker/> and [https://ec.europa.eu/jrc/sites/default/files/20150930-second-best-practices-tto-circle-gentile\\_en.pdf](https://ec.europa.eu/jrc/sites/default/files/20150930-second-best-practices-tto-circle-gentile_en.pdf)

<sup>32</sup> <https://creativecommons.org/2021/05/24/were-turning-20-whats-happened-since-2001/>

<sup>33</sup> <https://oaspa.org/information-resources/frequently-asked-questions/#FAQ1>

<sup>34</sup> For example, Fact Sheets by Creative Commons UK on Open Science and on Open Access both acknowledge that using a CC BY-NC license will not qualify as Open Access as defined in the Berlin Declaration, Bethesda Statement on Open Access Publishing, and Budapest Open Access Initiative (accessible at <https://doi.org/10.5281/zenodo.841086> and <https://doi.org/10.5281/zenodo.840652>). Additionally, Creative Commons has acknowledged that only its CC BY and CC BY-SA licenses are conformant with the Open Definition by the Open Knowledge Foundation (<https://creativecommons.org/2014/10/07/open-definition-2-0-released/>).

<sup>35</sup> Guidance on open science and licensing of research products by the Ecological Society of America is typical in this regard (<https://www.esa.org/openscience/sample-page-2/open-science-licenses/>).

<sup>36</sup> <https://sparcopen.org/open-data/>

conforms to the “full Open Definition of Open Data” and suggests using those proposed by Creative Commons (CC).<sup>37</sup> As noted previously, CC licenses, however, span a broad spectrum of conditions for use, not all of which permit reuse “for any other purpose.” Despite this incongruity, both the definition of open data and the use of open standard licenses are relevant to consider in the DSI discussion.

Indeed, standard open licenses are increasingly being used to facilitate access to data in online repositories thanks to widespread adoption of the FAIR Data Principles<sup>38</sup>, which propose to make data findable, accessible, interoperable and reusable. FAIR principles recommend the use of an appropriate open license to provide legal certainty regarding reuse (considered further in Section 4 below in the context of EU policy).

The Open Knowledge Foundation has also developed standard open licenses as part of an Open Data Commons toolkit to assist users to publish, in publishing, providing and using open data.<sup>39</sup>

Open licenses have proved useful for standardizing degrees of openness and improving legal certainty concerning reuse of content to which they are applied. They are based on copyright, an intellectual property right that grants the creator of an original creative work the exclusive right to control the reproduction and modification of the work. Open licenses do not replace copyright but work alongside it, allowing authors to keep their copyrights while permitting others to use their works. However, questions<sup>40</sup> have been raised regarding the suitability of open licenses for research data, which may not always meet the minimum criteria for copyright protection in a creative work. This renders the permissions and restrictions embodied in an open license unenforceable as a content creator cannot enforce rights they do not have. Furthermore, research data may be protected by sui generis database rights.<sup>41</sup>

As seen in open source software licenses, standard open licenses also raise compatibility issues. Even among the relatively small set of Creative Commons licenses, substantial incompatibilities do exist, which can be particularly problematic for data reuse.<sup>42</sup> For these reasons, a number of organizations recommend using the CC0 waiver to share research data in order to remove legal uncertainty and maximize interoperability.<sup>43</sup> Challenges related to compatibility and

---

<sup>37</sup> Ibid.

<sup>38</sup> Wilkinson et al., “The FAIR Guiding Principles for Scientific Data Management and Stewardship.”

<sup>39</sup> <https://opendatacommons.org/index.html>

<sup>40</sup> As noted in Seitz, “Digital Sequence Information—Legal Questions for Patent, Copyright, Trade Secret Protection and Sharing of Genomic Sequencing Data.” See also Lawson and Rourke, “Open Access DNA, RNA and Amino Acid Sequences.”

<sup>41</sup> In certain jurisdictions such as the European Union, sui generis database rights may also exist that prevent copying and reusing of substantial parts of a database. Unlike copyright, database rights protect the maker’s investment, not originality. The European Directive 96/9/EC on the legal protection of databases which came into force on 1 January 1998 provides a database broadly as “a collection of independent works, data or other materials which are arranged in a systematic or methodical way and are individually accessible by electronic or other means.” Collections of data made or used in the course of research are generally understood to constitute a database subject to legal protection (see, for example, the University of Reading’s guidance considering database rights in research data under UK law: <https://www.reading.ac.uk/research-services/research-data-management/data-management-planning/intellectual-property-rights-and-research-data>).

<sup>42</sup> For a visual summary see [https://commons.wikimedia.org/wiki/File:Vectorized\\_CC\\_License\\_Compatibility\\_Chart.svg](https://commons.wikimedia.org/wiki/File:Vectorized_CC_License_Compatibility_Chart.svg)

<sup>43</sup> See [https://wiki.creativecommons.org/wiki/CC0\\_use\\_for\\_data](https://wiki.creativecommons.org/wiki/CC0_use_for_data). As a pertinent example, Figshare, a web-based interface designed for academic research data management and research data dissemination adopts a CC0 waiver as the default tool for researchers to share their datasets. Its Copyright and License Policy notes that CC0 waiver is used by hundreds of organizations and its rationale for using the CC0 waiver preferentially over open license options to disseminate research data, explicitly acknowledges that uncertainty exists as to whether highly factual data and databases are protected by

interoperability (even within a relatively small license suite of the Creative Commons) may caution against the development of standard open licenses based to enforce ABS-related rights.

The “full Open Definition of Open Data” by SPARC (see above) does permit some restrictions if they are “necessary to *preserve provenance and openness*.” The full Open Definition (version 2.1) provides comprehensive criteria for “open”, including for the data or content to be available in the public domain (defined as “the absence of copyright and similar restrictions, whether by default or waiver of all such conditions”). Alternatively, data can be made available under an “open license” that meets detailed criteria mandating use, redistribution, modification, and compilation for any purpose while preventing discrimination against any person or group (e.g. commercial users) and also mandates the right to make derivatives.

With regard to open data, two other influential community declarations bear mentioning. The first is the 2010 Panton Principles for Open Data in Science.<sup>44</sup> They note that for data to be effectively used and added to by others, data should be made open in accordance with the Open Definition: freely available on the public internet permitting any user to download, copy, analyze, re-process, pass them to software or use them for any other purpose without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. It cautions that many widely recognized licenses are not intended for, and are not appropriate for, data or collections of data. It strongly discourages against the use of licenses that limit commercial reuse or the production of derivative works by excluding use for particular purposes or by specific persons or organizations. The second is the Principles of Open Scholarly Infrastructure (POSI), which address the organization, sustainability and governance of open infrastructure, i.e. databases and other services that support open science.<sup>45</sup> They endorse open source software licensing, open data using the CC0 waiver as best practice subject to privacy and data protection laws, patent non-assertion commitments for open infrastructure, and provide further clarification on what it means to make data “available” as a large infrastructure provider, recommending, for example, the release of periodic “data dumps.”

## Open DSI and ABS

---

### DSI and public databases

Depositing genetic sequence data in public databases has become the standard scientific practice in life science research and innovation. Owing to the plummeting costs of DNA sequencing over the past 30 years, over 238 million high-quality genetic sequences (and over 1.5 billion reads) are currently available through the synchronized databases of the International Nucleotide Database Collaboration (INSDC).<sup>46</sup> The INSDC constitutes a core infrastructure<sup>47</sup> used by the global

---

copyright or other rights and databases may contain facts that, in and of themselves, are not protected by copyright law (<https://help.figshare.com/article/copyright-and-license-policy>).

<sup>44</sup> Murray-Rust et al., “Panton Principles.”

<sup>45</sup> Bilder, Lin, and Neylon, “The Principles of Open Scholarly Infrastructure.”

<sup>46</sup> Comprising GenBank in the U.S.A., the European Nucleotide Archive (ENA) in the United Kingdom, and the DNA Data Bank of Japan (DDBJ). For more see <http://www.insdc.org/> and Arita et al., “The International Nucleotide Sequence Database Collaboration.”

<sup>47</sup> A 2019 analysis of more than 2,600 public biological databases found that 99% of all biological databases that provide access to nucleotide sequence data were directly or indirectly dependent on the INSDC database infrastructure (Rohden et al., “Combined Study on Digital Sequence Information in Public and Private Databases and Traceability.”).

scientific community to efficiently store, share, analyze and link<sup>48</sup> nucleotide sequence data and other data associated with genetic resources across a vast public biological database ecosystem. With a combined annual cost of approximately US\$ 50-60 million, borne predominantly by the governments of the United States, Japan, and countries of the EU,<sup>49</sup> the INSDC helps maintain a vast data commons as a global public good. The ubiquitous availability of genetic sequence and related data has fueled the growth of life sciences research and innovation and is a key enabler of the emerging field of synthetic biology.

## DSI and ABS policy tensions

In the past decade, however, tensions have emerged between “open access” DSI, and the “fair and equitable benefit-sharing” obligations arising from sovereign rights enshrined in international treaties that govern access to, and sharing of benefits arising from, the utilization of genetic resources (“ABS”).<sup>50</sup> Critics contend that open access to DSI undermines ABS by, in effect, providing access to information on genetic resources decoupled from the “fair and equitable benefit-sharing” obligations triggered by their utilization in research and development. They are concerned that because DSI is a crucial enabler of life sciences research, innovation and the emerging synthetic biology market, continued open access to DSI without concomitant benefit-sharing will only further undermine and entrench inequities in ABS implementation.

## What can the INSDC’s access policy tell us about open access for DSI?

INSDC’s access policy does not include any reference to “open access” per se. Instead, the policy refers to “free and unrestricted access” to DSI and specifically prohibits any restrictions, licensing requirements or licensing fees in relation to the use or redistribution of any specific sequence or the database.<sup>51</sup>

Why is the INSDC relevant? Because INSDC is the core infrastructure for nucleotide sequence data that links to other biological data across all branches of the life sciences. In short, it is a critical starting point of the scientific record. INSDC’s access policy is the prevailing model for accessing DSI across public databases. Furthermore, INSDC is inter-connected with 2,000 downstream databases that “inherit” INSDC’s “free and unrestricted access” terms when they access nucleotide sequence data from the core infrastructure.

The INSDC’s access policy is typical of the access policies of most -omics and biological datasets comprising DSI. INSDC’s access policy goes back to data sharing practices that evolved during

---

<sup>48</sup> For example, Accession Numbers (comprising unique and permanent identifiers issued by the INSDC for each submitted sequence) link sequence data with the biological samples and associated metadata from which they are derived.

<sup>49</sup> The European Nucleotide Archive (ENA), administered by European Molecular Biology Laboratory (EMBL), is funded predominantly, but not exclusively, by EU member states. Australia, for example has been an associate member since 2008 and Argentina was an associate member from 2014 to 2020. See <https://www.embl.org/about/member-states/>. For more on the funding estimate see Rohden et al., “Combined Study on Digital Sequence Information in Public and Private Databases and Traceability.”

<sup>50</sup> Primarily under the Convention on Biological Diversity and its Nagoya Protocol, and extending to other United Nations processes, such as the FAO International Treaty on Plant Genetic Resources for Food and Agriculture (ITPGRFA), the WHO Pandemic Influenza Preparedness Framework (PIP) and marine biodiversity of areas beyond national jurisdiction (BBNJ) under the UN Convention on the Law of the Sea (UNCLOS).

<sup>51</sup> INSDC Policy, approved 2012, available at <http://www.insdc.org/policy.html>. See also Brunak et al., “Nucleotide Sequence Database Policies.”

the Human Genome Project (HGP) in which the Bermuda Principles<sup>52</sup> agreed in 1996 committed scientists involved in the HGP to make sequence assemblies and their annotations freely available in the public domain for both research and development in order to maximize benefits to society. Subsequently, the 2003 Fort Lauderdale agreement<sup>53</sup> and the 2009 Toronto agreement extended the concept beyond the HGP to include pre-publication access to other -omics and biological datasets.<sup>54</sup> The requirement to publish nucleotide sequence data is intended to enable scientific reproducibility, perpetuate scientific integrity, and accelerate further research and innovation.

INSDC's access policy and the open and reusable<sup>55</sup> data sharing practices that have emerged in relation to DSI raise two questions. First, whether "open access" means or is equivalent to "free and unrestricted access"? Second, whether access terms that are not "free and unrestricted" (for example, restricted access, licensing requirements or licensing fees) are inconsistent with open access?

Critics argue the current DSI data commons should not be maintained at the expense of ABS and that irrespective of the INSDC's disclaimer, "free and unrestricted access" does not mean (or should not be understood to mean) "free of ABS obligations". This reasoning finds support in INSDC's constituent databases. Although GenBank's terms of use repeat the "free and unrestricted" language used by the INSDC, it also acknowledges that restrictions such as "other intellectual property rights" claimed by "submitters of the original data (or the country of origin of such data)" may apply. It notes that it "cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in the molecular databases."<sup>56</sup> Similarly, EMBL-EBI's terms of use "places no additional restrictions on the use or redistribution of scientific data available via its online services." However, it acknowledges that such data "may be subject to rights claimed by third parties, including but not limited to, patent, copyright, other intellectual property rights, biodiversity-related access and benefit-sharing rights."<sup>57</sup>

INSDC's public comments are also illustrative. It notes "that INSDC has been long recognized as the reliable framework for sustainably maintaining nucleotide sequence data and associated metadata throughout the scientific community [and this] framework became a model for the Open Access movement for academic literature." It explicitly cautions that its access policy "should not be misinterpreted as open access, where use restrictions may apply depending on the licensing terms." In other words, its policy does not impose any restriction on use as per its "free and

---

<sup>52</sup> 1996 Bermuda Meeting Report available at <https://dukespace.lib.duke.edu/dspace/handle/10161/7715>.

<sup>53</sup> The Wellcome Trust 2003. Sharing Data from Large-scale Biological Research Projects: A System of Tripartite Responsibility. Fort Lauderdale, USA.

<sup>54</sup> Rohden et al., "Combined Study on Digital Sequence Information in Public and Private Databases and Traceability."

<sup>55</sup> This trend has not been entirely without consequence in the scientific community as is evident from contrasting perspectives published in recent years in *Science* (see for example, Amann et al., "Toward Unrestricted Use of Public Genomic Data"; Pennisi, "Group Calls for Rapid Release of More Genomics Data"; Nicol et al., "Consent Insufficient for Data Release.").

<sup>56</sup> As per the terms of use determined by the US National Center for Biotechnology Information (which administers Genbank) accessible at <https://www.ncbi.nlm.nih.gov/home/about/policies/>. The text reads "NCBI itself places no restrictions on the use or distribution of the data contained therein. Nor do we accept data when the submitter has requested restrictions on reuse or redistribution. However, some submitters of the original data (or the country of origin of such data) may claim patent, copyright, or other intellectual property rights in all or a portion of the data (that has been submitted). NCBI is not in a position to assess the validity of such claims and since there is no transfer of rights from submitters to NCBI, NCBI has no rights to transfer to a third party. Therefore, NCBI cannot provide comment or unrestricted permission concerning the use, copying, or distribution of the information contained in the molecular databases."

<sup>57</sup> EMBL-EBI's terms of use are accessible at <https://www.ebi.ac.uk/about/terms-of-use/>

unrestricted” policy. Yet, it also acknowledges its reliance on sister databases to govern access to human genomes, which are subject to access restrictions in order to comply with informed consent terms or requirements under law or ethical frameworks in research.<sup>58</sup>

What can we extrapolate from this? INSDC’s policy and public comments reinforce the notion that different types of open access can exist, in which “free and unrestricted” is one type of open access that differs from other open access models in which re-use can be restricted in some ways. GenBank’s and EMBL-EBI’s terms of use demonstrate that “free and unrestricted” open access is not guaranteed given the overriding disclaimers regarding third-party rights. Therefore, although such access terms do not preclude the imposition of benefit-sharing-related obligations, the existence of any such obligations (e.g. pursuant to national legislation with implications for the access and use of DSI) are not immediately apparent to a user as the INSDC databases does not have a means of communicating or transmitting such terms. By extrapolation across the DSI database ecosystem, it ultimately falls on the user to determine their freedom to operate for any uses they may make of these databases and the data and information derived from those databases<sup>59</sup> In practice, however, many users operate under the assumption that freedom to operate is assured and that DSI is open access, available in a free and unrestricted manner.

Overall, INSDC’s policy does show that DSI are currently overwhelmingly open and globally available. It allows users to access and use the entire dataset across the public database infrastructure under standardized terms of use without explicit restrictions, permissions or constraints on reuse. This can be distinguished from the experience in open access publishing, which is characterized by significant fragmentation. If the access modality for biological data in public databases were to be mapped on the open access spectrum, as depicted in Table 2, it would be closely aligned with the Open Knowledge’s definition of open data and occupy the least restrictive and most-permissive form of open access-libre available on the spectrum. The only identifiable restriction is that third party rights are not abdicated.

The distinction between open access publishing and open access to DSI can be attributed to their different histories. Whereas open access publishing arose to counter a prevailing closed subscription model for scientific journals, the public database infrastructure for biological data was born open, precisely to facilitate widespread data sharing and use. Their distinct lineages are evident in the iterative public declarations, which have sought to provide guidance and consensus on what constitutes “open” in each context.

This distinction between the degrees of openness that characterizes open access publishing and the standardized approach to open access that characterizes biological data<sup>60</sup> has profound consequences. Beyond immediate data-access implications, the degree of interoperability that results from a lack of fragmentation has profound technical and technological implications for the database infrastructure and service interoperability. Important technological tools for scientists such as API-REST services, data transmission protocols, and automated updates all rely on

---

<sup>58</sup> Arita et al., “The International Nucleotide Sequence Database Collaboration.”

<sup>59</sup> The legal burden and risk borne by users of the INSDC databases concerning freedom to operate is also noted by Lawson et al. “Open Access DNA, RNA and Amino Acid Sequences: The Consequences and Solutions for the International Regulation of Access and Benefit-sharing.”

<sup>60</sup> In which, except for limited exceptions related to human genetic data, as a result of legitimate ethical and privacy concerns, the entire database is accessible on identical terms.

cohesive and standardized conditions. The degree of interoperability is the most defining characteristic of open access in the context of public DSI, as this allows the global scientific community to efficiently store, share, analyze, and link data across a vast public biological database ecosystem.

## Can controlled and/or differentiated access provide an alternative to open access to DSI?

---

As noted in the introduction, although there appears to be convergence on the importance of open access to DSI in general, divergent political views have been voiced. Parties have questioned whether open access should be free, restricted or unrestricted, regulated or unregulated, subject to free, prior and informed consent (from some groups, such as indigenous peoples and local communities, or from all), or not permitted at all.<sup>61</sup> This raises the question as to whether alternative models for accessing DSI are available.

In Table 1, we consider existing models of controlled and differentiated access to biological data in order to evaluate whether they are capable of achieving benefit-sharing policy objectives for DSI as described above (i.e. ensuring access without significantly impairing research and innovation on the one hand and achieving benefit-sharing on the other). What is evident from existing controlled access initiatives is that they are not able to provide the interoperability that characterizes the current system. The lack of interoperability can be caused by a need to negotiate to obtain ex-ante approvals to access and use for specific purposes (i.e. the status quo for genetic resources, but not, in most instances, for DSI). Or caused at the database or dataset level wherein differentiated terms of use and conditions lead to compatibility issues that result in fragmentation akin to open source and Creative Commons licensing. The ensuing disruptions to frictionless data sharing and interoperability of DSI that would be caused by wide-scale implementation of the models in Table 1 would likely significantly hinder research and innovation in the life sciences.

Turning to models of open but differentiated access in a DSI context, similar concerns arise. The Global Biodiversity Information Facility (GBIF),<sup>62</sup> which provides “free and open access to biodiversity data”, is an illustrative example of the interoperability challenges that arise from differentiated access models due to license compatibility issues, even when standard open licenses are used. GBIF’s 62,657 datasets are available under CC0, CC BY and CC BY-NC license options. However, the 7,303 datasets available<sup>63</sup> under the CC BY-NC license (nearly 12% of the database) are not available for commercial use. This means the dataset as a whole is not available for certain types of research and innovation. Assuming it would likely be very difficult to negotiate bilateral permissions for the data that is associated with certain types of licenses, it is likely that users would simply exclude this subset of data from their research. In other words, licenses ultimately enable a simple method for jurisdiction shopping by excluding those data with unattractive conditions.

---

<sup>61</sup> See Tshitwamulomoni and Voigt-Hanssen, “Co-Leads’ Summary of the Discussion of the Contact Group Regarding Areas of Potential Convergence and of Apparent Divergence on Digital Sequence Information on Genetic Resources.”

<sup>62</sup> An international network and data infrastructure coordinated through a Secretariat in Copenhagen, which is funded by the world’s governments and is aimed at providing anyone, anywhere, open access to data about all types of life on Earth (<https://www.gbif.org/>).

<sup>63</sup> As at 20 October 2021.

**Table 1.** Comparison of controlled access models and their implications for scientific research and innovation

Controlled access example	Brief description	Access Modality and restrictions on use	Current ABS compatibility	Potential to hinder scientific research and innovation if applied to DSI at scale <sup>64</sup>
<p>GISAID  <a href="https://nbnatlas.org/">https://nbnatlas.org/</a></p>	<p>A global science initiative that provides access to genomic data of influenza viruses and the coronavirus responsible for the COVID-19 pandemic</p>	<p>Controlled access via online registration and acceptance of standard terms of use.<sup>65</sup> Use is subject to attribution obligations, transfers to unregistered users and <b>use of sequences in other databases is prohibited.</b></p>	<p>Low (does not address ABS obligations, objective is not to promote monetary benefit-sharing)</p>	<p>High (data are siloed and is not interoperable until/unless permission provided)</p>
<p>European Genome-phenome Archive (EGA)  <a href="https://ega-archive.org/">https://ega-archive.org/</a></p>	<p>EGA is a permanent archive and sharing hub for all types of personally identifiable human genetic and phenotypic data resulting from biomedical research</p>	<p>Hybrid model. Differentiated access reflects an “as open as possible as closed as necessary” approach in which user access and obligations are governed by Data Access Agreements (DAA), which are annotated with standardized “consent codes.” Datasets which are subject to access controls require a user to apply to a Data Access Committee (DAC) and negotiate a DAA based on a template.<sup>66</sup></p>	<p>Low (consent codes do not address ABS obligations and its objective is not to promote monetary benefit-sharing)</p>	<p>High (consent code fragmentation<sup>67</sup> and custom DAAs raise compatibility issues which likely reduce interoperability, and certain uses require bilateral permission).</p>
<p>Finnish Biodiversity Information Facility (FinBIF)  <a href="https://laji.fi/en">https://laji.fi/en</a></p>	<p>A (mostly) open access portal receives, stores and manages biodiversity data mobilized in Finland</p>	<p>Hybrid model. Open access repository with controlled access for sensitive data. FinBIF stores controlled-access data and intermediates user access requests with owners.</p>	<p>Medium (access to controlled data needs to be negotiated with data owner on a case by case basis).<sup>68</sup></p>	<p>High (restricted access data are siloed and is not interoperable until/unless bilateral permission is provided)</p>

<sup>64</sup> i.e. if applied as the prevailing access model to all biological and related data in public databases which are derived from genetic resources under sovereign control

<sup>65</sup> See the GISAID EpiFlu™ Database Access Agreement available at <https://www.gisaid.org/registration/terms-of-use/>

<sup>66</sup> The archives terms of use (<https://ega-archive.org/data-use-conditions>), which lists the consent codes, notes that EGA have implemented the Data Use Ontology (DUO) maintained at <https://github.com/EBISPOT/DUO> and first described in Dyke et al., “Consent Codes.”

<sup>67</sup> The consent codes include: DUO:0000004 no restriction; DUO\_0000042 General Research Use; DUO:0000006 health/medical/biomedical research and clinical care; DUO:0000007 disease-specific research and clinical care; DUO:0000011 population origins or ancestry research; DUO:0000012 research-specific restrictions; DUO:0000015 no general methods research; DUO:0000016 genetic studies only; DUO:0000018 not-for-profit use only; DUO:0000019 publication required; DUO:0000020 collaboration required; DUO:0000021 ethics approval required; DUO:0000022 geographical restriction; DUO:0000024 publication moratorium; DUO:0000025 time limit on use; DUO:0000026 user-specific restriction; DUO:0000027 project-specific restriction; DUO:0000028 institution-specific restriction; DUO:0000029 return to database/resource.

<sup>68</sup> This creates potential for avoidance and jurisdiction shopping for least-restrictive terms.

<p>GA4GH Beacon project <a href="https://beacon-project.io/">https://beacon-project.io/</a></p>	<p>Developed under an initiative by the Global Alliance for Genomics and Health (GA4GH) for the federated discovery of genomic data in biomedical research and clinical applications.</p>	<p>Implements ELIXIR AAI for genomic data sharing with a three-tier access system (public, registered and controlled) that controls what kind of information will be provided to different types of users.<sup>69</sup></p>	<p>Medium (access to controlled data is passport and permissions based and the terms of use could be designed to address ABS).</p>	<p>High (restricted access data permissions likely to impose compatibility issues, reduce interoperability and require bilateral negotiations for certain uses)</p>
<p>NBN atlas <a href="https://nbnatlas.org/">https://nbnatlas.org/</a></p>	<p>Comprises the UK's largest collection of freely available biodiversity data and is maintained by the National Biodiversity Network (NBN)</p>	<p>Hybrid model. Datasets are available under a number of standard options, primarily Creative Commons licenses. Data records for species listed on the Sensitive Species List are blurred to the resolution stated in the list (i.e. data are not published in full). Sensitive records at a higher resolution are subject to a tailored Restricted Re-use Licence issued by the relevant Data Partner.<sup>70</sup></p>	<p>Medium (access to controlled data could be subject to terms of use designed to address ABS but only for species listed on the Sensitive Species List)<sup>71</sup></p>	<p>High (differentiated access terms, whether customized or standard, are likely to impose compatibility issues that reduce interoperability and customized access terms will require bilateral negotiations for certain uses)</p>

Models of open but differentiated access terms restricting commercial use, whether on the basis of user type or intention, has implications for the frictionless interoperability of DSI. Furthermore, all models of differentiated access inherently create the potential for avoidance and jurisdiction shopping for least-restrictive terms, which should be a key consideration in the design of any DSI benefit-sharing mechanism as it would have the potential to undermine benefit-sharing objectives.

Whether or not controlled access mechanisms or even standard open licenses could be designed specifically to address ABS concerns related to DSI is a separate question that is not considered in this paper. However, given the inherent interoperability implications of these models, if applied at scale, these options may face significant challenges.

## What do international, regional, and national policies tell us about open access for DSI?

The shift in international science moving towards greater openness and data sharing is evident in the UNESCO Recommendation on Open Science. It is also reflected in the European Union's approach and by national governments in different regions that increasingly promote open data and open science, not just as a catalyst to research, innovation and technological development, but as

<sup>69</sup> ELIXIR Authentication and Authorisation Infrastructure (AAI) provides controlled access via online permissioning and passports, which allows service providers (both in academia and industry) to control and manage access rights of their users and create different access levels. The public tier, open to all users, will only disclose allele frequencies in the genomic data; the registered users (scientists working at universities and research centres) have access to cohort specific allele frequencies; users in the controlled tier have to be individually approved by Data Access Committees and have access to individual level data from cohorts.

<sup>70</sup> As per the terms governing use of the NBN Atlas websites: <https://docs.nbnatlas.org/nbn-atlas-terms-of-use/>

<sup>71</sup> This creates potential for avoidance and jurisdiction shopping for least-restrictive terms if data from species that are not on the Sensitive Species List can be used as an alternative, or if data from a species on the list can be accessed on more favourable terms than data related to another species on the list.

integral features of policies governing innovation and sustainable innovation economic development. We will look at the status in Europe in detail, followed by a high-level overview of science policy and innovation in the African Union, India and China. Beyond national and regional science and innovation policy, the policies of scientific journals and public funders mandating the deposit of biological data in public databases demonstrate a widespread trend towards increased openness in research data.

## UNESCO

“Open” approaches in research and innovation contexts appear to be gaining significant traction internationally. This is best exemplified by the 40th session of UNESCO’s General Conference in 2019. Its 194 Member States tasked the Organization with developing an international standard-setting instrument on Open Science. UNESCO’s General Conference adopted the UNESCO Recommendation on Open Science in November 2021.<sup>72</sup> The Recommendation provides a framework to support scientific cooperation and seeks to leverage and amplify the existing open ecosystem to make science more transparent, accessible, equitable and inclusive. It includes definitions for a number of open elements, including open science, open scientific knowledge and open research data, in which timely, free and open access to research data – subject to access restrictions where proportionate and justified – is considered a foundational pillar of open science.

“Open research data” is explicitly characterized as data that can be *“openly used, reused, retained and redistributed by anyone, subject to acknowledgement.... available in a timely and user-friendly, human and machine readable and actionable format in accordance with principles of good data governance and stewardship, notably the FAIR (Findable, Accessible, Interoperable, and Reusable) principles.”* Interestingly, the Recommendation includes overarching principles and shared values, which touch on themes of diversity, inclusion and transparency as being integral dimensions of open science. The Recommendation is particularly relevant to the current discussion concerning open access in the context of DSI as 98% of Parties to the CBD are also Member States of UNESCO.<sup>73</sup> Additionally, unlike the various statements, principles and declarations, which are mentioned throughout this brief as addressing “openness” over the past two decades, adoption of the Recommendation represents an international consensus on various open elements associated with open science, including principles and guidelines for their interpretation and implementation.

## European Union

The principles of open access publications, set forth in the Budapest Open Access Initiative (2002), the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities (2003) and the Bethesda Statement (2003) have been the foundation for the current open access landscape across the EU. Concrete measures to achieve the principles proposed therein, through the self-archiving (i.e. green route) and open-access journals (i.e. gold route) possibilities have further institutionalized these principles. Furthermore, the proactive approach by the academic community by engaging and endorsing open access efforts, such as the Open Access Initiative (OA2020

---

<sup>72</sup> “UNESCO Recommendation on Open Science - UNESCO Digital Library.”

<sup>73</sup> i.e. with the exception of the European Union, Israel, and Liechtenstein all Parties to the CBD are also UNESCO Member States, although Switzerland has established a National Organizing Committee (NOC) for Liechtenstein as a dependent territory. See <https://www.cbd.int/information/parties.shtml> and <https://en.unesco.org/countries> for lists of parties to and members of the CBD and UNESCO, respectively.

initiative, 2016)<sup>74</sup> and cOAlition S (2018)<sup>75</sup>, to transition into a full open access environment has been key to ensure realization. The OA2020 initiative mainly focuses on reallocating the journal subscription funds into sustainable open access models that would accelerate the transition to a long-term open access environment. The goal of the cOAlition S alliance is the implementation of Plan S<sup>76</sup>, which has the target of ensuring that research results financed by its public and private funding bodies be immediately and fully published in open access journals or platforms (i.e. gold option) or repositories (i.e. green option) by 2021. It acknowledges, however, that the transition period will include some ‘hybrid’ options or longer waiting times for book and monograph formats.

Interestingly, Plan S defines an open access journal as one where peer-reviewed research articles are openly available, but other content may remain behind a paywall. However, such content likely comprises highly curated or “added value” such as interactive tools or graphics, videos, etc. and not the “raw data” currently available through public databases per-se. Directive (EU) 2019/1024 on open data and the reuse of public sector information (PSI) was adopted on 20 June 2019 for implementation by Member States by 16 July 2021. It builds on a legal framework established in 2003, which seeks to establish an EU open data market as a key building block of the overall EU data economy.<sup>77</sup>

This Directive underpins the EU Research and Innovation Strategy 2020-2024 that seeks to empower and revitalize the European Research Area, including, as one of its six priorities, “access to and transfer of scientific knowledge including knowledge circulation and open access, international cooperation.”<sup>78</sup> Since the PSI Directive governs data resulting from public funding, it also directly informs the EU’s open research data policy, which requires that data should be openly accessible and reusable.

The EU’s research strategy and data policy are implemented through ‘Horizon Europe’, the EU’s research and innovation programme, which has a budget of EUR 95.5 billion for the period 2021-2027. Horizon Europe implements an open science policy that mandates immediate open access to all scientific publications and responsible research data management so that data are Findable, Accessible, Interoperable and Re-usable (FAIR).<sup>79</sup> Horizon Europe, therefore distinguishes between open access to scientific peer-reviewed publications and open access to research data, which instead is required to be made ‘as open as possible, as closed as necessary’ in order to safeguard legitimate interests or constraints.<sup>80</sup>

Tailored guidance in this regard has yet to be developed for Horizon Europe. However, guidance under its predecessor framework programme, Horizon 2020 (2014-2020), encouraged recipients to maximize access to and reuse of research data. Researchers should take into account the need to balance openness and protection of scientific information, commercialization and Intellectual

---

<sup>74</sup> <https://oa2020.org/be-informed/#about>

<sup>75</sup> <https://www.coalition-s.org/about/>

<sup>76</sup> <https://www.coalition-s.org/why-plan-s/>

<sup>77</sup> Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information. Subsequently revised by Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information.

<sup>78</sup> [https://ec.europa.eu/info/sites/default/files/rtd\\_sp\\_2020\\_2024\\_en.pdf](https://ec.europa.eu/info/sites/default/files/rtd_sp_2020_2024_en.pdf)

<sup>79</sup> Directorate-General for Research and Innovation (European Commission) 2021

<sup>80</sup> H2020 Programme Guidelines on FAIR Data Management in Horizon 2020. Available at [https://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-data-mgt\\_en.pdf](https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf).

Property Rights (IPR), privacy concerns, security as well as data management and preservation questions.<sup>81</sup>

Under Horizon 2020, this approach was coupled with a requirement for a Data Management Plan (for every project funded) that identifies in advance whether data will be shared/made open access and explains legal, contractual, ethical or voluntary considerations that may restrict open access to research data. Additionally, the submission of Horizon 2020 proposals involved a self-assessment to evaluate compliance with ethical rules and standards, relevant European legislation, international conventions and declarations, national authorizations and ethics approvals, and social impact of their planned research, including, where appropriate, the EU ABS Regulation, in which case the project is subject to reporting requirements concerning due diligence. Presently the EU ABS Regulation does not require due diligence for the use of DSI from public databases.

## African Union

In June 2014, the African Union approved a Science, Technology and Innovation Strategy for Africa 2024 (STISA-2024),<sup>82</sup> a 10-year strategy approved as part of the long-term African Union Agenda 2063, which is underpinned by science, technology and innovation as multi-function tools and enablers for achieving continental development goals. It acknowledges that a multi-disciplinary and multi-sectoral approach driven by open data, open innovation and entrepreneurship are essential to achieving a knowledge economy and sustainable socio-economic development across Africa.

Entrenching the open science movement is suggested by some scholars as a powerful means to put Africa on a path to sustainable development. However studies of Africa's open science landscape reveal significant challenges in data management capacities across the region.<sup>83</sup> As has been noted in the ABS policy discussions, actors who lack the capacity to use DSI cannot benefit from it, even if it is, in theory, accessible.<sup>84</sup> This can be extrapolated to apply to open research data more generally and highlights the need to strengthen capacity development and technology transfer in Africa in order to unlock the potential of open science as a path to sustainable development. Only limited resources have been made available for this purpose through the CBD to date. Furthermore, it is not clear whether capacity development and technology transfer alone will sufficiently address the "gap" between the capacities of actors in developing and developed countries.<sup>85</sup>

## Brazil

Following the introduction of a federal law in 2016 designed to foster scientific development, research, scientific and technological capacity-building and innovation<sup>86</sup>, Brazil's Ministry of Science, Technology, Innovations and Communications (MCTIC) launched a National Innovation Policy in 2020. This creates an Inter-ministerial Committee that has been tasked with formulating a National Innovation Strategy. Initial proposals focus on expanding research infrastructure, streamlining the patenting process and encouraging open scientific knowledge available on digital

---

<sup>81</sup> Ibid.

<sup>82</sup> Available at [https://au.int/sites/default/files/newsevents/workingdocuments/33178-wd-stisa-english\\_-\\_final.pdf](https://au.int/sites/default/files/newsevents/workingdocuments/33178-wd-stisa-english_-_final.pdf).

<sup>83</sup> Mwelwa et al., "Developing Open Science in Africa."

<sup>84</sup> Karger, du Plessis, and Meyer, "Digital Sequence Information on Genetic Resources (DSI): An Introductory Guide for African Policymakers and Stakeholders."

<sup>85</sup> Ibid.

<sup>86</sup> Marco Legal de Ciência, Tecnologia e Inovação - Lei 13.243, de 11 de janeiro de 2016.

platforms as a means to accelerate efforts in science, technology and entrepreneurship for a more harmonious relationship with the planet.

During this period, Brazil's ABS framework – which was first introduced in 2000 and since inception has applied to genetic information as well as genetic resources with both comprising “genetic heritage” – has undergone a significant transformation. In 2016, Brazil transitioned from an individualized bilateral ABS system requiring PIC and MAT to a standardized bilateral ABS system that operates on a declaratory basis requiring users of genetic resources and information to declare activities in an online registration system, SisGen. During the 15-year period in which the individual PIC and MAT system was in place, 2,600 access authorizations and 295 benefit-sharing agreements were processed. By contrast, 47,000 access activities and 1,500 benefit-sharing arrangements were processed during the first 1.5 years of the standardized registration system, in which 449 and 64, respectively, were related to genetic information related to genetic heritage from in silico origin.<sup>87</sup>

## India

In October 2020, India's Ministry of Science & Technology released a draft National Science, Technology and Innovation Policy that is guided by the vision of positioning India among the top three scientific superpowers in the coming decade. Open science, including open access to publications and open data resulting from publicly-funded research, is proposed as a central feature of the policy in order to foster more equitable participation in science through increased access to research outputs, greater transparency and accountability in research. It proposes that minimal restrictions better promote resource utilization through a constant exchange of knowledge between producers and users of knowledge. The policy also proposes ‘One Nation, One Subscription and Democratization’, which envisions free access to all journals, Indian and foreign, under a centrally-negotiated payment mechanism that seeks to democratize science by providing access to scholarly knowledge to not just researchers but also to every individual in the country.<sup>88</sup>

## China

China's transition to an innovation-based economy is clearly evident in its latest Five Year Plan (2021–2025), which provides a strategic, innovation-driven blueprint for Chinese development in the short- and medium-term. China's Minister of Science and Technology, Wang Zhigang, has indicated that a 15-year strategy is being developed for science and technology development that will significantly boost spending in basic and applied research and lay the groundwork for long-range objectives for national economic and social development through the year 2035. There are reports that as part of this strategy, China is working on a master plan for the internationalization of its domestic journals and plans to pursue an open science strategy at a national level.<sup>89</sup>

This gradual shift towards open science is evident, for example, in predictions in 2019 that “in 10 years’ time we will see the realization of almost complete and immediate open access to publicly

---

<sup>87</sup> See “Brazil's Position on DSI” submitted in response to CBD Notification 2019-012 <https://www.cbd.int/abs/DSI-views/2019/Brazil-DSI.pdf>.

<sup>88</sup> [https://dst.gov.in/sites/default/files/STIP\\_Doc\\_1.4\\_Dec2020.pdf](https://dst.gov.in/sites/default/files/STIP_Doc_1.4_Dec2020.pdf)

<sup>89</sup> As indicated in the opening address by the Executive Secretary of the China Association for Science and Technology at the 43rd Annual Meeting of the Society of Scholarly Publishing (SSP) held in May 2021 (see “China ‘Pursuing National Open Science Strategy.’”)

funded research, especially for journal articles, because of the push for innovation-driven development for the Chinese economy and society.”<sup>90</sup> Similarly, a dedicated program, launched in March 2020 by the China Association for Science and Technology<sup>91</sup> funded studies on global trends in Open Access and the development approach of Open Access in China.<sup>92</sup>

---

<sup>90</sup> International Science Council, “Open Access in China: Interview with Xiaolin Zhang of the National Science Library.”

<sup>91</sup> A non-profit, non-governmental organization of Chinese scientists and engineers, which is composed of 167 national professional societies and hundreds of local branches at various levels.

<sup>92</sup> Ning and Zhao, “To Embrace Open Science More Closely.”

**Table 2.** Unpacking the open access spectrum in a range of contexts.

	<b>Closed access</b>	<b>Charged access</b>	<b>Controlled access</b>	<b>Open access (gratis)</b>	<b>Open access (libre)</b>			<b>Public domain</b>
<b>Permissions or freedoms</b>				freedom from price barriers only	freedom from price barriers and one or more permission barriers			freedom from price and all permission barriers
<b>Use or reuse</b>				free to read (no or uncertain rights of reuse)	free to reuse (with degrees of reuse)			free to reuse without any restriction
<b>Consistent with BBB OA publishing?</b>	not within scope				within scope			
<b>Consistent with Open Definition?</b>	not within scope					within scope		
<b>Consistent with FAIR Data Principles?</b>	As open as possible and as closed as necessary					spans the full spectrum of closed to open		
<b>Consistent with Panton Principles?</b>	not within scope						within scope	
<b>Conformant licenses or equivalent</b>		© All rights reserved		No license or terms of use that do not confer FTO	CC BY-NC; CC BY-ND; CC BY-NC-ND	CC BY (most permissive); CC BY-SA	INSDC Terms of Use	CC0 (no rights reserved waiver)
<b>Relevant Examples</b>	Proprietary information managed confidentially	Journal subscriptions	Finish Biodiversity Initiative; Global Alliance for Genomics & Health; ELIXIR AAI; GISAID				INSDC databases & >2000 downstream databases	Dryad research data repository

## Conclusion

---

As we have seen, the research communities that produce DSI are characterized by a strong tradition of open sharing, and the INSDC's policy of "free and unrestricted access" is both a product of this tradition, and a key enabler of research and innovation in these communities. Existing practices in these fields lean toward being fully open when we consider openness on a spectrum. For an overview of this spectrum, Table 2 maps various elements noted in this paper.

Interoperability is a defining characteristic of access to biological data in public databases. It permits frictionless use of DSI across a vast database and software infrastructure, by humans and algorithms alike. Data interoperability is maximized, and friction is minimized by near uniform terms of access and conditions on use that, save for exceptions related to ethics and privacy concerns for human-sourced data, enable "unrestricted" data sharing and use of biological data. However, this does not necessarily mean such data are free of ABS obligations or other third-party rights, but for the user such constraints are effectively invisible. This frictionless data sharing and use minimizes transaction costs in accessing and using data if they were subject to specific permissions or restrictions on use, or if datasets or databases were accessible on differentiated and potentially incompatible terms.

This has implications for the design of benefit-sharing from DSI. First, to maintain the high-degree of interoperability that characterizes the status quo, a multilateral access model applied as universally (i.e. across all DSI and all international benefit-sharing fora) as possible should be favored. This could take the form of uniform terms of access for DSI across public databases. Benefit-sharing obligations that apply to the entire DSI dataset globally will best protect the open system because the overall conditions change, but the technological infrastructure, which enables the DSI ecosystem to function and generate knowledge, would not require massive changes. These types of benefit-sharing obligations are decoupled from access, which remains open. In comparison, options that require accounting of DSI access, movement, and use (bilateral mechanisms) appear more likely to impair interoperability and to generate high transactions costs and frictions to data flow that will significantly hinder research.

A historical context for open access and open scientific research data can serve as a starting point for characterizing and defining open access in the context of DSI. This may assist in developing a working definition tailored to DSI. However, a consensus definition might not necessarily need to be the primary focus. Instead, we advise that policy discussions should pay closer attention as to whether, and to what extent, scientific research and innovation would be significantly hindered by changes to the current "open and unrestricted" access and use of DSI in public databases. This lens appears better suited to guide discussions on the design of the access pillar for any potential benefit-sharing solution.

Taking inspiration from SPARC's approach, it may be useful to look beyond the deceptively simple question, "what is open access?" towards a more nuanced approach for designing and evaluating ABS policy solutions, for example, by asking, "is it as open as possible?" Efforts should be made to ensure that any changes made to the status quo, if necessary, are proportionate and justified. Applying this nuanced approach to benefit-sharing objectives suggest that a multilateral and universal mechanism for DSI would appear to constitute "as open as possible" provided it can be

designed to deliver benefits that are deemed acceptable by Parties. Certainly, the scientific community's quest to ensure that open access to DSI will continue to be guaranteed, that biological data will be publishable, available, linkable, downloadable, and continue to flow into the downstream databases and software that they use every day, is strongly aligned with international, regional and national policies concerning science and innovation. In short, the world appears to be moving towards greater openness to promote research, innovation and technological development with the goal of sustainable economic development. The outstanding question is whether the CBD and the broader benefit-sharing community will support or go against this trend and what the consequences of that choice will be.

## References

---

- Amann, Rudolf I., Shakuntala Baichoo, Benjamin J. Blencowe, Peer Bork, Mark Borodovsky, Cath Brooksbank, Patrick S. G. Chain, et al. "Toward Unrestricted Use of Public Genomic Data." *Science* 363, no. 6425 (January 25, 2019): 350–52. <https://doi.org/10.1126/science.aaw1280>.
- Arita, Masanori, Ilene Karsch-Mizrachi, Guy Cochrane, and on behalf of the International Nucleotide Sequence Database Collaboration. "The International Nucleotide Sequence Database Collaboration." *Nucleic Acids Research* 49, no. D1 (January 8, 2021): D121–24. <https://doi.org/10.1093/nar/gkaa967>.
- Bauer, Bruno, Guido Blechl, Christoph Bock, Patrick Danowski, Andreas Ferus, Anton Graschopf, Thomas König, et al. "Recommendations for the Transition to Open Access in Austria," November 30, 2015. <https://doi.org/10.5281/zenodo.34079>.
- Bilder, Geoffrey, Jennifer Lin, and Cameron Neylon. "The Principles of Open Scholarly Infrastructure," 2020. <https://doi.org/10.24343/C34W2H>.
- Brunak, Soren, Antoine Danchin, Masahira Hattori, Haruki Nakamura, Kazuo Shinozaki, Tara Matisse, and Daphne Preuss. "Nucleotide Sequence Database Policies." *Science* 298, no. 5597 (November 15, 2002): 1333–1333. <https://doi.org/10.1126/science.298.5597.1333b>.
- "China 'Pursuing National Open Science Strategy.'" *Research Information*, June 3, 2021. <https://www.researchinformation.info/news/china-pursuing-national-open-science-strategy>.
- Conference of the Parties to the Convention on Biological Diversity. "14/20. Digital Sequence Information on Genetic Resources." Decision. Convention on Biological Diversity, November 30, 2018. <https://www.cbd.int/doc/decisions/cop-14/cop-14-dec-20-en.pdf>.
- Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information, Pub. L. No. 32003L0098, OJ L 345 (2003). <http://data.europa.eu/eli/dir/2003/98/oj/eng>.
- Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information, Pub. L. No. 32013L0037, OJ L 175 (2013). <http://data.europa.eu/eli/dir/2013/37/oj/eng>.
- Dyke, Stephanie O. M., Anthony A. Philippakis, Jordi Rambla De Argila, Dina N. Paltoo, Erin S. Luetkemeier, Bartha M. Knoppers, Anthony J. Brookes, et al. "Consent Codes: Upholding Standard Data Use Conditions." *PLOS Genetics* 12, no. 1 (January 21, 2016): e1005772. <https://doi.org/10.1371/journal.pgen.1005772>.
- International Science Council. "Open Access in China: Interview with Xiaolin Zhang of the National Science Library," February 25, 2019. <https://council.science/current/blog/open-access-in-china-interview-with-xiaolin-zhang-of-the-national-science-library/>.
- Karger, Elizabeth, Pierre du Plessis, and Hartmut Meyer. "Digital Sequence Information on Genetic Resources (DSI): An Introductory Guide for African Policymakers and Stakeholders." The ABS Capacity Development Initiative, 2019. <https://unctad.org/system/files/official-document/ditc-ted-05052020-BioTradeSSC-DSI.pdf>.
- Lawson, Charles, and Michelle Rourke. "Open Access DNA, RNA and Amino Acid Sequences: The Consequences and Solutions for the International Regulation of Access and Benefit Sharing." SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, October 5, 2016. <https://doi.org/10.2139/ssrn.2848136>.
- MarketsandMarkets. "Synthetic Biology Market by Tools (Oligonucleotides, Enzymes, Synthetic Cells), Technology (Gene Synthesis, Genome Engineering), Application (Tissue Regeneration, Biofuel, Consumer Care, Food & Agriculture, Environmental) and Region - Global Forecast to 2026," May 2021. <https://www.marketsandmarkets.com/Market-Reports/synthetic-biology-market-889.html>.
- Moore, Samuel A. "A Genealogy of Open Access: Negotiations between Openness and Access to Research." *Revue Française Des Sciences de l'information et de La Communication*, no. 11 (August 1, 2017). <https://doi.org/10.4000/rfsic.3220>.
- Murray-Rust, Peter, Cameron Neylon, Rufus Pollock, and John Wilbanks. "Panton Principles." Panton Principles, February 19, 2010. <http://pantonprinciples.org/>.
- Mwelwa, Joseph, Geoffrey Boulton, Joseph Muliaro Wafula, and Cheikh Loucoubar. "Developing Open Science in Africa: Barriers, Solutions and Opportunities." *Data Science Journal* 19, no.

- 1 (August 5, 2020): 31. <https://doi.org/10.5334/dsj-2020-031>.
- Nicol, Dianne, Lisa Eckstein, Heidi Beate Bentzen, Pascal Borry, Mike Burgess, Wylie Burke, Don Chalmers, et al. “Consent Insufficient for Data Release.” *Science* 364, no. 6439 (May 3, 2019): 445–46. <https://doi.org/10.1126/science.aax0892>.
- Ning, Bi, and Yan Zhao. “To Embrace Open Science More Closely.” *The Innovation* 1, no. 1 (May 21, 2020). <https://doi.org/10.1016/j.xinn.2020.04.012>.
- Open-ended Working Group on the Post-2020 Global Biodiversity Framework. “Digital Sequence Information on Genetic Resources,” July 5, 2021. <https://www.cbd.int/doc/c/707f/6565/1b79a3df3aea669e19078615/wg2020-03-04-en.pdf>.
- Pennisi, Elizabeth. “Group Calls for Rapid Release of More Genomics Data.” *Science*, May 22, 2009. <https://doi.org/10.1126/science.324.1000b>.
- Publications Office of the European Union. *The Economic Impact of Open Data: Opportunities for Value Creation in Europe*. LU: Publications Office, 2020. <https://data.europa.eu/doi/10.2830/63132>.
- Rohden, Fabian, Sixing Huang, Gabriele Dröge, and Amber H. Scholz. “Combined Study on Digital Sequence Information in Public and Private Databases and Traceability.” Convention on Biological Diversity, January 31, 2020. <https://www.cbd.int/doc/c/1f8f/d793/57cb114ca40cb6468f479584/dsi-ahteg-2020-01-04-en.pdf>.
- Sara, Rodrigo, Andrew Lee Hufton, and Amber Hartman Scholz. “Compatible or Incompatible? DSI, Open Access, and Benefit-Sharing.” Preprint. SocArXiv, December 22, 2021. <https://doi.org/10.31235/osf.io/nw8g9>.
- Seitz, C. “Digital Sequence Information—Legal Questions for Patent, Copyright, Trade Secret Protection and Sharing of Genomic Sequencing Data.” *IOP Conference Series: Earth and Environmental Science* 482, no. 1 (March 1, 2020): 012002. <https://doi.org/10.1088/1755-1315/482/1/012002>.
- Suber, Peter. “Gratis and Libre Open Access.” *SPARC Open Access Newsletter*, 2008. <https://dash.harvard.edu/handle/1/4322580>.
- Tennant, Jonathan P., François Waldner, Damien C. Jacques, Paola Masuzzo, Lauren B. Collister, and Chris H. J. Hartgerink. “The Academic, Economic and Societal Impacts of Open Access: An Evidence-Based Review.” F1000Research, September 21, 2016. <https://doi.org/10.12688/f1000research.8460.3>.
- Tshitwamulomoni, Lactitia, and Gaute Voigt-Hanssen. “Co-Leads’ Summary of the Discussion of the Contact Group Regarding Areas of Potential Convergence and of Apparent Divergence on Digital Sequence Information on Genetic Resources.” Convention on Biological Diversity, September 1, 2021. <https://www.cbd.int/doc/c/a7d7/f5cf/c99a6073666521fafa5b320b/wg2020-03-cg-05-report-en.pdf>.
- “UNESCO Recommendation on Open Science - UNESCO Digital Library.” Accessed December 21, 2021. <https://unesdoc.unesco.org/ark:/48223/pf0000379949.locale=en>.
- Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3, no. 1 (March 15, 2016): 160018. <https://doi.org/10.1038/sdata.2016.18>.
- Woelfle, Michael, Piero Olliaro, and Matthew H. Todd. “Open Science Is a Research Accelerator.” *Nature Chemistry* 3, no. 10 (October 2011): 745–48. <https://doi.org/10.1038/nchem.1149>.