Open Access to Digital Sequence Information Benefits the Three Objectives of the Convention on Biological Diversity

Joint Authors:

Ruth Bastow, Ph.D. The Earlham Institute, UK (Global Plant Council Committee Member)
Richard Bruskiewich, Ph.D. Dept. of Botany, Faculty of Science, University of British Columbia
Carolyn Lawrence-Dill, Ph.D. Dept. of Genetics, Development, and Cell Biology and Dept. of Agronomy, Iowa State University, Ames, IA, USA
Shawn Dorius, Ph.D. Dept of Sociology, Iowa State University, Ames, IA, USA
Paul Kersey, Ph.D. EMBL-The European Bioinformatics Institute, UK
Emily Marden, J.D. Faculty of Law, University of British Columbia, Canada
Nicola Patron, Ph.D. The Earlham Institute, UK
Ruaraidh Sackville Hamilton, Ph.D. International Rice Research Institute, Los Baños, Laguna, Philippines
Norman Warthmann, Ph.D. Departments of Plant Science and Ecology, Evolution, and Genetics, The Australian National University, Canberra, Australia

Abstract

Article 17.1 of the Convention on Biological Diversity (CBD) places an obligation on Contracting Parties to facilitate the exchange of information relevant to the conservation and sustainable use of biological diversity. We argue that Digital Sequence Information is relevant to these purposes, and that the objectives of the CBD are best served by encouraging the generation of digital sequence information on genetic resources and fostering the open sharing of this information. By contrast, the restriction of access to sequence information will impede global benefit sharing and the objectives of the CBD.

The Convention on Biological Diversity

The objectives of the Convention of Biological Diversity (CBD) are the conservation of biological diversity, the sustainable use of its components, and the fair and equitable sharing of benefits arising out of the use of genetic resources.

The CBD is clear that scientific information relevant to conservation and sustainable use should be shared. Article 17.1 states, "[t]he Contracting Parties shall facilitate the exchange of information, from all publicly available sources, relevant to the conservation and sustainable use of biological diversity." In addition, and relevant to this obligation, Article 12 states that Contracting Parties shall:

- Establish and maintain programmes for scientific and technical education and training in measures for the identification, conservation and sustainable use of biological diversity and its components and provide support for such education and training for the specific needs of developing countries;

- Promote and encourage research which contributes to the conservation and sustainable use of biological diversity, particularly in developing countries; and
- Promote and cooperate in the use of scientific advances in biological diversity research in developing methods for conservation and sustainable use of biological resources.

The directives aim at inclusive sharing of scientific research by and between all Contracting Parties: “Each Contracting Party shall endeavour to develop and carry out scientific research based on genetic resources provided by other Contract Parties with the full participation of, and where possible in, such Contracting Parties.” CBD Article 15.6.

The importance of shared research to support the goals of the CBD is also recognised by Article 8 of the Nagoya Protocol to the CBD, which states that each party shall: “Create conditions to promote and encourage research which contributes to the conservation and sustainable use of biological diversity, particularly in developing countries, including through simplified measures on access for non-commercial research purposes ...”

Benefits of Open Access to DSI

Over recent decades, the ability to assay biological systems has been transformed by the development of new technologies. Researchers can now easily generate and analyze sequence information from polymeric molecules such as DNA, RNA and proteins present in cells and organisms. Early techniques to generate this information required extensive human involvement, and as a result, was typically generated in small scale and published alongside its interpretation in scientific articles. Newer sensor and assay technologies with lower cost and higher throughput have since emerged. The information they generate is necessarily captured, stored, analysed, and disseminated in electronic, digital form. This **digital sequence information (DSI)** has already had a significant impact on research and development in many fields, including biomedicine, animal and plant breeding, and biodiversity conservation. In particular, it is now inexpensive to quickly produce genome sequence information for large numbers of individual samples, which enables new approaches for the characterisation, understanding, and sustainable use of biological entities and their interactions.

Consistent with the CBD “aim of sharing in a fair and equitable way the results of research and development and the benefits … arising from the utilization of genetic resources” (Article 15.7), we outline four benefits that arise from **open access** to DSI.

First, a single sequence has less value than a collection of sequences. This is true both in terms of relating different measurements from the same biological system (for example, transcriptomic or proteomic data is understood in the context of a genome); but also when comparing within or between populations and species. The essential principle of molecular biology is that conservation of sequence implies conservation of function, and conversely, that the origin of divergent biology lies in sequence variation. The ability to access and compare sequences from multiple samples provides an opportunity for insight into the function and novelty of differences.

As a sequence is refined, annotated with descriptive information, and made relational with other sequences, its value and potential for use is increased. Shared data can be quality-controlled and (re-)sequenced to create aggregate sequence assemblies which show genetic variants. These can be annotated and linked to relevant metadata that add further value to the DSI. This process of annotating and comparing sequence assemblies ultimately yields insights on differences that play a role in function. Open access to DSI empowers continued third party annotation after initial public release in a manner that benefits all users; closing off access deprives all users of these additional benefits.

Second, in contrast to physical materials, where sharing often diminishes value (i.e., the more broadly a physical material is shared, the further it must be subdivided) the value of digital information is **increased** by sharing.[1] The use of digital information does not diminish its availability or value. In fact, the more it is shared, the more broadly its benefits can be used. Indeed, the collection of huge sets of data, stored, analysed and shared in digitized, electronic form, allows the use of statistical analysis techniques capable of finding patterns (and the absence of patterns) and thereby generating and testing scientific hypotheses. The separation of data into individual silos would likely contribute to a 'tragedy of the anti-commons', in which fencing-off of a resources diminishes both the value of individual data items and the number of people who can access them, reducing the potential for the productive use of the resource.[2]

Third, shared data can minimize duplication of efforts, improve research reproducibility, facilitate the establishment of scientific collaborations beyond institutional and international boundaries, and help to combat scientific fraud. It can also be argued that original generators of DSI do not truly "own" the data unless the data are shared; without sharing, public attribution of "first across the line" cannot be defensibly asserted and a second or third group could be attributed with DSI characterization of an organism, thus ultimately staking the claim of ownership.

Fourth, open data can widen the research community, becoming part of a common infrastructure encompassing both data, computational and biotechnological tools, which a diverse range of new actors, including entrepreneurs, citizen scientists, and local researchers can draw on as capital to challenge the monopolies of powerful incumbents. This process is analogous to the current global democratization of information technology brought about by the expansion of open source software and affordable access to inexpensive computing devices and performant internet connectivity, in particular, in the developing world (e.g. see example below on Bioinformatic Capacity Building in Africa).

Open Access to DSI Benefits the Three Objectives of the Convention

Open DSI supports the stated aims of the CBD. Below we outline several examples where open access and sharing of DSI have specifically enabled the goals of the CBD to be met. These examples provide evidence of the benefits to broad and unimpeded access and availability to

[1] Borgman, C. L. (2012), The conundrum of sharing research data. J *Am Soc Inf Sci Tec* 63: 1059–1078..
[2] Heller, Michael A. (1998) The Tragedy of the Anticommons: Property in the Transition from Marx to Markets, *Harv. L. Rev.* 111: 621-88.

DSI by the global community. **Limiting or discouraging the generation or sharing of DSI would impede the application of these highly successful models to other problems**.

1. Conservation

The CBD defines "*in situ* conservation" as "the conservation of ecosystems and natural habitats and the maintenance and recovery of viable population of species in their natural surroundings and, in the case of domesticated or cultivated species, in the surroundings where they have developed their distinctive properties." Consistent with the CBD goal of conservation, the United Nation's Sustainable Development Goal (SDG) 2 ("End World Hunger"), urges that we "maintain the genetic diversity of seeds, cultivated plants and farmed and domesticated animals and their related wild species".

Raw sequence data can be annotated with information on evolutionary conservation and variation (between individuals, populations or species) as well as with functional regions including, *inter alia* protein-coding sequences, regulatory regions, and regions packed into structural proteins. This component of DSI to is essential to interpret sequence data and use it to enhance conservation and sustainable use. Comparative studies of DSI of different species and of different individuals of the same species, commonly known as comparative genomics, is one of the primary methods for the identification of specific sequences responsible for traits, described by biologists as the linking of genotypes to phenotypes.

The conservation of biological diversity is critically dependent on inventory systems that collect, identify, curate, and ultimately promulgate biological diversity. It will not be possible to immediately halt and reverse the loss of biodiversity, but aggressive efforts to capture and digitize plant genetic material is critical to preserve what remains of biological diversity. DSI directly informs conservation by cataloguing the sequence variation that underpins biodiversity and is critical for conducting surveys of the genetic variability of populations of organisms. Efforts to document and digitize plant genetic material is essential for realizing the conservation objective of the CBD. At present, the lack of large-scale, interoperable data systems makes it impossible to a) quantify how much genetic variation currently exists in wild and cultivated species, and b) monitor progress on biodiversity.

These benefits are illustrated in the following examples:

NuNet - Global Collaboration to Identify Environmental Threats

Humans are impacting ecosystems through the combustion of fossil fuels, agricultural fertilization, introductions of invasive species, destruction of habitat, and selective hunting and fishing. However, there have been alarming few globally coordinated experiments to quantify the general impacts to ecological systems. The Nutrient Network (NutNet)[3] is a grassroots research effort to address these questions within a coordinated research network comprised of more than 40 grassland sites worldwide. The network collects samples and

[3] http://www.nutnet.org/home

data from multiple sites using identical protocols. These data, including DSI, are shared and compared to ask general questions like, 'What is controlling diversity and productivity?', 'How are human activities changing diversity?' and 'How will these changes impact the environment further on down the road?' This globally collaborative, data-sharing approach has allowed them to answer questions on the effects of fertilization on the diversity and stability of ecosystems (it decreases diversity) and on the role sunlight and herbivores play in controlling diversity of grassland plants (herbivore grazing allows light to reach more species)[4]. Because of the large amount of data from diverse global environments, the findings can be applied to inform conservation practices in a wide number of countries.

Responding to the Ash Dieback outbreak

In mid-September 2012, conservation volunteers recognized that ash trees (the third most common tree in Britain) in an ancient woodland in Norfolk, UK were showing unusual disease symptoms. Following analysis of the diseased material, DNA sequencing confirmed that the infection was a fungal pathogen spreading across Europe. Recognizing that Norfolk was likely at the edge of the epidemic spreading from Europe, scientists at several research institutions formed an *ad hoc* emergency response group to start sequencing the genomes of the pathogen and infected trees. In the hope of encouraging a rapid response, they agreed to make all data immediately available with Open Access. They built websites and software tools to make data available and accessible to genomic scientists, biologists and the public, for who they built a Facebook game that received over 63,000 plays in the first year enabling the public to be involved in the rapid assembly of the genomes. The result of the 'Open Dieback Project' was an unprecedented speed of discovery. In just a few years the consortium was able to identify genetic markers for trees with low susceptibility to the disease and the fundamentals required to select and breed trees with enhanced tolerance to the disease.[5]

2. Sustainable Use

The CBD defines "sustainable use" as "the use of the components of biological diversity in a way and at a rate that does not lead to the long-term decline of biological diversity, thereby maintaining its potential to meet the needs and aspirations of present and future generations."

Sustainable use of biological systems requires approaches that go beyond political boundaries, since biological systems, by their nature, do not respect such boundaries. Meeting global challenges relating to sustainable use, including crop disease epidemics, alien species invasions, habitat loss, and climate change adaptation, are better tackled with complete knowledge of the diversity and population structure.

DSI has numerous benefits for sustainable use, as illustrated in the following examples:

[4] Grace et al (2016) *Nature* 529, 390–393.
[5] NORNEX Consortium Final Report, available at http://oadb.tsl.ac.uk/wp-content/uploads/2016/04/Nornex_Final_Report_April_2016.pdf.

Understanding complex microbial ecosystems

Most members of the microbial communities e.g. of the gut, skin, soil and oceans belong to phyla from which no isolates have been cultured owing to their unknown growth requirements. This ignorance is problematic as alterations in these population and their ecology have implications for human health, crop yields and the environment. DSI from these microbial communities is now being captured and catalogued in metagenomics projects where all sequence information (of all species) in a sample from a specific environment is collected. This information can help to identify native, invasive and endangered species and track their dynamics. As an example, the microbial communities which inhabit soils are some of the most complex but remain poorly understood despite their economic importance in crop production including fixing atmospheric nitrogen, nutrient cycling and sequestration, and suppressing diseases. Functional metagenomics strategies are being used to explore the interactions between plants and microbes through cultivation-independent study of microbial communities providing insights into the role of previously uncultivated microbes in the promotion of plant growth. The success of such investigations relies on access to DSI from the broadest number of possible environments. This will inform the maintenance and creation of healthy soils for sustainable productivity and ecosystem health.

Sustainable Bioproduction of Taxol for Chemotherapies

The effort to develop a low-cost, sustainable source of paclitaxel provides a good example how DSI can support sustainable use. Paclitaxel, currently sold under the brand name Taxol, is a chemotherapy medication used to treat a number of types of cancer. Discovered in 1971 in the bark of a tree, Pacific Yew (*Taxus brevifolia*), and approved for medical use in 1993, it is now recognized by the WHO as one of the most effective and safe medicines needed in a healthcare system. Clinical trials in the mid 1980s required the felling of thousands of trees and led to the source trees becoming an endangered species. Since then paclitaxel has been produced via semi-synthetic chemistry and plant cell culture. These advances have taken pressure off the Pacific Yew, reducing the threat of species loss. However, supply and price vary, and the drug frequently features on the Drug Shortages List compiled by the American Society of Health System Pharmacists[6]. The biosynthetic pathway has now been mined from the *T. brevifolia* genome and, because these data are shared in public databases, a large number of international research groups are able to use these DSI to work towards the monumentally difficult task of reprogramming species that are suitable for low-cost, large-scale production (e.g., yeast, bacteria) with the large and complex paclitaxel biosynthetic pathway. As well as providing the potential for meeting global demands through sustainable, low-cost production, this method has the potential to enable bioproduction of novel chemical derivatives of the natural product – e.g., with reduced side-effects.

[6] Elzawawy et al (2013) Variation in the availability of cancer drug generics in the United States of America, Annals of Oncology 24, Suppl 5, v17-v22.

Improving the Sustainability of Agriculture

Up to 40% of crop losses are due to pests and diseases. A significant route to improving the sustainability of agriculture is to limit the agrichemicals used to kills pests and pathogens by improving the disease resistance of crops. Plants carry heritable resistance genes that protect them against species strains of diseases; improving agriculture involves identifying the origins of strains of pathogens so that cognate resistance genes can be found and bred into field varieties to provide resistance. In 2016, Bangladesh's wheat crop suffered from an outbreak of an aggressive fungus known as wheat blast. In some regions, losses were up to 70%. To rapidly determine the precise identity and likely origin of the outbreak pathogen, scientists applied field pathogenomics (transcriptome sequencing of symptomatic and asymptomatic leaf samples collected from infected wheat fields in Bangladesh). To encourage more experts to use their resources and expertise to find a solution, all raw sequence data was immediately and openly released on the project's website[7]. Phylogenomic and population genomic analyses revealed that the Bangladesh wheat blast outbreak was likely caused by isolates belonging to the South American wheat-infecting lineage of *M. oryzae*. Data sharing drew together an international group of experts, who compared global data on blast pathogens and identified that the new Bangladesh pathogen was most likely introduced into Asia from South America[8] (fungal pathogens are known to be transported across contents by wind, as well as by trade). The result was that, in record time, the knowledge acquired to manage wheat blast in Brazil using disease resistant cultivars and fungicides can be directly applied to treat a Bangladeshi epidemic. This highlights the need for intensive monitoring and surveillance of crop pathogens and also the power of making DSI open and accessible for rapidly addressing sustainability of food production.

3. Fair and Equitable Sharing of Benefits

In increasing scientific knowledge of biodiversity, open access to DSI supports the benefits identified in the Nagoya Protocol, the SDGs identified by the CBD as having particular relevance,[9] and certain of the Aichi Biodiversity Targets.[10]

Examples of these benefits include potential impacts on response to disease outbreaks and impacts for global food security:

Monitoring and responding to outbreaks of infectious diseases

[7] Open Wheat Blast - http://www.wheatblast.net

[8] Islam et al (2017) BMC Biology 14:84 https://doi.org/10.1186/s12915-016-0309-7.

[9] https://sustainabledevelopment.un.org/?menu=1300.These include SDGs 2.4, 2.5, 6.6, 8.4, 12.8, 14.2 and 15.9.

[10] https://www.cbd.int/sp/targets/. These include Strategic Goal B (reduce the direct pressures on biodiversity and promote sustainable use), Strategic Goal C (improve the status of biodiversity by safeguarding ecosystems, species and genetic diversity), Strategic Goal D (enhance the benefits to all from biodiversity and ecosystem services), and Strategic Goal E (Enhance implementation through participatory planning, knowledge management and capacity building).

The first full genome sequence for a human bacterial pathogen, *Haemophilus influenzae*, was completed in 1995. Since then, the development of sequencing technologies has made genomic analysis of emerging pathogens easier, faster, and less expensive; instead of taking months or weeks, such investigations can often be accomplished in days. For example, within days of the initial identification of the first cases of 2009 pandemic influenza A (H1N1) in spring 2009, scientists had identified the origin of all eight influenza virus gene segments. Within two weeks, the US Center for Disease Control and Prevention (CDC) began to distribute RT-PCR diagnostic test kits to public health laboratories. To build capacity for rapid responses to such outbreaks, in 2012, the United States Defense Advanced Research Projects Agency (DARPA) issued a challenge for rapid production of influenza vaccine[11]. A Canadian company, Medicago, was able to respond, and produced 10 million doses of the vaccine in a wild relative of tobacco within one month of receipt of an emailed genetic sequence. This ability to utilize DSI to respond quickly is not limited to human disease; for crops, a novel approach called "field pathogenomics" has recently been implemented for pathogen population surveillance. This method, based on sequencing technology, allows scientists to acquire data directly from field samples of pathogens, warning growers about new races of disease emerging on previously resistant varieties[12,13].

Improvement of plants for global food security

All crop improvement practices aim to capture genetic variants that confer desirable traits. Historically, genetic variation was recognised and tracked through visual assessment of variation. Discovery was serendipitous and fixation took a long time; centuries to millennia in some cases[14]. With current genomic analysis tools it is straightforward and cost-effective to ascertain and accurately quantify the level of variation within and between crops. This is used to inform quality control, maintenance, distribution, and use of ex-situ collections in seed banks. When linked to measured traits, digital sequence information allows repositories to be searched for genetic resources that likely exhibit desired characteristics, greatly accelerating our progress towards new, more resilient and more diverse crops that will be key for the sustainable intensification of our agricultural systems. Foregoing the step of linking phenotypes to causal individual loci, Genomic Selection is being used to realise major performance gains in livestock, and it has similar potential in plants. In either case, accurate prediction is enhanced not only by access to more data, but to more variation in the data, which means it is important to sample the gene pool as extensively as possible, and to assay under the widest range of environmental conditions, which will require global collaborative efforts, for which data sharing is instrumental and digital sequence information

[11] Lomonossoff and D'Aoust (2016) Plant-produced biopharmaceuticals: A case of technical developments driving clinical deployment. *Science* 353(6305): 1237-1240.
[12] https://www.jic.ac.uk/news-and-events/news/2016/03/field-pathogenomics-cereal-killers/
[13] Meyer et al (2017) Quantifying airborne dispersal routes of pathogens over continents to safeguard global wheat supply. *Nature Plants* 329:1786.
[14] Fuller DQ. Contrasting Patterns in Crop Domestication and Domestication Rates: Recent Archaeobotanical Insights from the Old World. Ann Bot. Oxford University Press; 2007 Oct 1;100(5):903–24.

will be key to mainstream agricultural research[15]. Shared Digital Sequence Information has already contribute to greatly reducing the number of required breeding cycles, but we need to accelerate progress much more to meet current and future challenges.

Article 17.1 of the CBD requires that Contracting Parties' actions to facilitate the exchange of information take into account the **special needs of developing countries**. In the context of DSI, there is a "digital divide" that must be closed to enable those in developing countries to gain the full benefits of DSI and to allow for use of DSI to address local problems. Work on closing the digital divide is well underway:

Building Bioinformatics Capacity in Africa

Over the last decade, several international organisations[16] and better-resourced African establishments (e.g., The African Society of Human Genetics and the African Society for Bioinformatics and Computational Biology) have focused efforts on building capacity in bioinformatics. The popularity of bioinformatics is due to its versatility and infrastructure requirements and these efforts have been incredibly successful. For example, there is now an extensive Pan African Bioinformatics network, H3ABioNet[17], comprising 32 bioinformatics research groups distributed amongst 15 African countries[18]. In 2014, researchers in Kenya and South Africa led the sequencing and genome assembly of the tsetse fly, the vector of human African trypanosomiasis[19]. Nigerian bioinformatics research groups have applied bioinformatics techniques to a number of domestic issues, including malaria[20], while Ghanaian bioinformaticians have contributed to and led projects to analyse the sequence diversity of a wide range of human and plant pathogens and crop species[21]. These represent a considerable and rapidly-expanding knowledge base in genomics and bioinformatics, poised to maximise the use of DSI and other bioinformatic resources. Without access to open sequence databases and software, these knowledge exchanges and capacity building exercises would not have been possible. Limiting access to DSI would certainly curtail the expansion of these programs that rely on international collaborators working together on large and Open DSI datasets.

Fast and Frugal Lab Tools for Developing Countries

The synthetic biology community aims to make software, hardware and wetware for biology and biotechnology cheaper, easier to operate and internationally compatible. This is partly to facilitate scaling for industry, but this so-called "frugal science" movement also aims to

[15] Warthmann, N (2014) Plant Genetic Resources for Food and Agriculture and Genomics: Mainstreaming Agricultural Research through Genomics, ITPGRFA/FAO, Rome, www.fao.org/3/a-be655e.pdf.
[16] http://planetearthinstitute.org.uk/the-genomics-revolution-building-bioinformatics-capacity-in-africa
[17] http://h3abionet.org
[18] http://h3abionet.org
[19] International Glossina Genome Initiative. Genome sequence of the tsetse fly (Glossina morsitans): vector of African trypanosomiasis. Science. 2014; 344: 380–386. pmid:24763584
[20] Fatumo et al (2014) Computational Biology and Bioinformatics in Nigeria. *PLoS Comput Biol* 10(4): e1003516. https://doi.org/10.1371/journal.pcbi.1003516,
[21] Karikari TK (2015) Bioinformatics in Africa: The Rise of Ghana? *PLoS Comput Biol* 11(9): e1004308. https://doi.org/10.1371/journal.pcbi.1004308.

make open, cheap and easy tools specifically to facilitate the adoption of biotechnologies that utilise DSI in developing countries. Efforts in the development of low-cost, open-source laboratory hardware have been particularly successful[22] and include molecular biology essentials, e.g., OpenPCR[23] and Biropette[24], the latter deployed in a range of workshops across several countries by TReND (Teaching and Research in (Neuro)science for Development) in Africa[25]. Open source versions of microscopes (e.g., The FlyPi[26] and The Waterscope[27]) and even robotics e.g., OpenTrons[28] are enabling researchers in less-resourced institutions to thrive. For example, community platforms such as Hackteria[29] enable researchers in developing countries to equip biology labs from OpenSource Hardware at less than 10% of the commercial price[30]. Similar efforts are underway to equip researchers with Open-Source wetware (e.g., The BioBricks Foundation[31] and OpenPlant[32]), including molecular (DNA-based) tools to stimulate innovation and entrepreneurship.

DNA synthesis allows for the utilisation of DSI without access to physical genetic materials. For example, several projects have been focused on adapting genetic sequence information for the expression of high value natural products in organisms that are amenable for large-scale, low-cost bioproduction, such as yeast. Well-known examples include the biosynthetic pathway for the antimalarial, artemisinin from *Artemisia annua*, vanillin from *Vanilla planifolia,* and nootkatone from *Citrus* × *paradisi* (grapefruit). While previously the requisite sequences may have been sourced and cloned from plant material, rapid reductions in the cost of *de novo* gene synthesis and innovation in biological engineering technologies means that new versions of sequences, specifically adapted for expression in new species, are typically designed and ordered from a commercial provider negating the need for access to physical materials. In the artemisinin, vanillin and nootkatone examples, the nature and source of the sequence information is known, making any monetary benefit sharing relatively straightforward. However, informed by available sequence information, synthetic biologists explicitly design novel sequences that will code for *novel molecules that are not known to exist in nature*. This method is being utilized both for the production of novel natural products as ingredients for industry (e.g., novel fragrances) and also to produce libraries of novel drug-candidates for functional

[22] Baden et al (2015) Open Labware: 3-D Printing Your Own Lab Equipment. *PLoS Biol* 13(3): e1002086. https://doi.org/10.1371/journal.pbio.1002086.
[23] http://openpcr.org
[24] https://open-labware.net/projects/biropettes/
[25] http://trendinafrica.org
[26] Chagas et al (2017) The €100 lab: A 3D-printable open-source platform for fluorescence microscopy, optogenetics, and accurate temperature control during behaviour of zebrafish, *Drosophila*, and *Caenorhabditis elegans*. *PLoS Biol* 15(7): e2002702. https://doi.org/10.1371/journal.pbio.2002702.
[27] http://www.waterscope.org
[28] http://opentrons.com
[29] http://www.hackteria.org
[30] Gibney (2016) Open Hardware pioneers push for low-cost lab kit. *Nature* 531, 147–148.
[31] https://biobricks.org
[32] Capacity Building for the Bioeconomy in Africa. An OpenPlant Report, https://static1.squarespace.com/static/54a6bdb7e4b08424e69c93a1/t/597bccff4402438918153c5c/1501285648350/Bakubung-FinalReport-Web.pdf.

screens. In this case, no individual sequence contributes more than another or anything specific. All sequences together serve as inspiration for a final, human-designed sequence, are equal.

Norms for Data Sharing

Ultimately, information sharing is not a new idea: the sharing of ideas has always been a driver of human progress. Sharing culture traces its roots to indigenous communities and traditional food production systems. Today, there are many models for information sharing, ranging from open publication of data as a public commons to publishing data under patent laws for licensing and commercial exploitation. Models for open access to DSI (and associated software for data analysis) are arguably better aligned to traditional norms of reciprocity and community than exclusionary access models.

Moreover, a culture of sharing has already been adopted by many producers and consumers of DSI in the global academic community, particularly among biological researchers. The development of sequencing technologies has been associated with the parallel development of open repositories for DSI, dating back over five decades: the Protein Data Bank, a repository of 3-D structural information on protein sequences, has been publicly searchable since 1971; the origins of the European Nucleotide Archive and GenBank (the two oldest collections of nucleotide sequences) date from 1982. The expectation among funding agencies, publishers of traditional scientific literature, and researchers themselves has been that deposition of data in such repositories is an obligation on producers of data; and that the repositories should provide access without charge to the global community of researchers.

In 1996, when international agencies were coordinating in the then-ambitious project to sequence a human genome, they agreed on the Bermuda Principles[33] to "make the entire sequence freely available in the public domain for both research and development in order to maximise benefits to society". Subsequently, the genomics research community has been at the forefront of advancing best practice for open data publication[34]. The success of these resources and associated scientific norms have helped lead the wider scientific community to recognise the desirability of maintaining publicly accessible data for the generation of knowledge and hypotheses[35], leading to the development of an "Open Science" movement in many scientific domains[36]; and the growth of new publishing models for scientific literature,

[33] Summary of Principles Agreed at the First International Strategy Meeting on Human Genome Sequencing, available at *http://www.casimir.org.uk/storyfiles/64.0.summary_of_bermuda_principles.pdf.*
[34] Editorial (2009) Prepublication data sharing, *Nature* 461, 168-170.
[35] Kaye et al (2009) Data sharing in genomics — re-shaping scientific practice, *Nature Reviews Genetics* 10, 331-335.
[36] See, e.g., GODAN (http://www.godan.info/pages/about-godan), Report of the National Academies (*The Role of Scientific and Technical Data and Information in the Public Domain: Proceedings of a Symposium*. Washington, DC: The National Academies Press. https://doi.org/10.17226/10785); EU Commission Recommendation on access to and preservation of scientific information (2012) (https://ec.europa.eu/digital-single-market/en/news/commission-recommendation-access-and-preservation-scientific-information).

whereby costs are covered by the authors, not by access charges.[37] The development of an increasing, open body of scientific information, especially in genomics, reflects the belief that longer-term societal benefits will be found from harvesting the whole, not from hoarding the parts; the model promises to be particularly advantageous to developing countries, which would otherwise stand to be priced out of access to knowledge by the already-wealthy world.

Finally, it is worth noting that UNICEF's principles for Innovation and Technology[38] in Development specifically promote the use of Open Standards, Open Data, Open Source Publications and the use of Creative Commons.

Disadvantage of Treating DSI as "Genetic Resources" under the CBD

Open access to DSI offers tangible benefits for the three aims of the CBD in a manner that would not be matched by seeking to apply the current model for access to and benefit sharing of physical genetic resources. Indeed, the current model mixes both open and closed access systems, with mixed results. Genetic resources are located in a varied assortment of public and private collections and fall under a patchwork of access mechanisms, some of which fall outside the scope of access regulated by the CBD[39]. The range of distinct systems for permitting access and use has allowed for extensive IP application, encouraged protectionist responses from select nations, and exploited gaps in local, national, and international laws and norms to further carve up the landscape of genetic resources.

Efforts to restrict or foreclose access to DSI would result in an analogous patchwork of mechanisms. Researchers currently operate in a data environment where they can readily access DSI with high environmental, scientific, and economic value from large (and growing) open-access data repositories such as Genbank and Gramene. Private data is largely inaccessible. Further Restrictions on access to DSI would result in higher transaction costs to generate additional DSI. The net result would likely be more extensive use of existing open-access DSI, and in the absence of a commons of data and supporting infrastructure, access would in practice be restricted to those sufficiently wealthy to develop an environment to enable analysis and exploitation. —Some take the position that access to DSI should be restricted so that the holder of that DSI can benefit if that material provides the basis of a 'blockbuster' drug. However, this model has been shown not to be generalisable to the use of biodiversity as a whole[40] which, in fact, requires a broad spectrum of data to generate any innovation, with the contribution of each individual datum hard to ascertain and only a small portion of the whole.

[37] Tennant et al (2016) The academic, economic and societal impacts of Open Access: an evidence-based review. *F1000Research* 5:632.
[38] UNICEF Principles for Innovation and Technology in Development, https://www.unicef.org/innovation/innovation_73239.html
[39] CBD Article 15.3: "For the purpose of this Convention, the genetic resources being provided by a Contracting Party, as referred to in this Article and Articles 16 and 19, are only those that are provided by Contracting Parties that are countries of origin of such resources or by the Parties that have acquired the genetic resources in accordance with this Convention"
[40] Jefferson et al (2015) Gene patent practice across plant and human genomes, *Nature Biotechnology* 33,1033–1038..