TEMPLATE FOR COMMENTS

Contact information	
Surname:	Prof. Dr. Jörg Overmann (spokesman of the Leibniz NP&DSI Ad Hoc Group)
Given Name:	
Government	
(if applicable):	
Organization:	**Leibniz Association** Chausseestraße 111 10115 Berlin Germany The Leibniz Association connects 91 independent research institutions throughout Germany that range in focus from the natural sciences, engineering, mathematics and environmental sciences via economics, spatial and social sciences to the humanities. It is one of the four major non-university research organisations in Germany. They are funded jointly by the Federal Government and the *Länder*, employing some 18,700 individuals, including 9,500 researchers. The entire budget of all the institutes is approximately 1.8 billion Euros. www.leibniz-association.eu The Leibniz NP&DSI Ad HOC Group were established by the member institutes of the Leibniz Research Alliance on Biodiversity (LVB) and the Section "Life Sciences", among them: - Bernhard Nocht Institute for Tropical Medicine, Hamburg - German Primate Center – Leibniz Institute for Primate Research, Göttingen - Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures, Braunschweig, Germany (spokesman) - Leibniz Institute for Zoo and Wildlife Research, Berlin - Leibniz Institute for Agricultural Engineering and Bioeconomy, Potsdam - Leibniz Institute For Farm Animal Biology, Dummerstorf - Leibniz-Institute of Freshwater Ecology and Inland Fisheries, Berlin - Leibniz Institute of Vegetable and Ornamental Crops, Großbeeren - Leibniz Institute for Baltic Sea Research, Warnemünde - Leibniz Institute of Plant Biochemistry, Halle - Leibniz Institute of Plant Genetics and Crop Plant Research, Gatersleben - Leibniz Centre for Tropical Marine Research, Bremen - Leibniz Centre for Agricultural Landscape Research, Müncheberg - Museum für Naturkunde Berlin - Leibniz Institute for Evolution and Biodiversity Science - Senckenberg Gesellschaft für Naturforschung (SGN), Frankfurt am Main - Zoological Research Museum Alexander Koenig, Bonn
E-mail:	Joerg.overmann@dsmz.de; amber.h.scholz@dsmz.de, premke@leibniz-gemeinschaft.de

Title of document reviewed:		The Emergence and Growth of Digital Sequence Information in Research and Development: Implications for the Conservation and Sustainable Use of Biodiversity, and Fair and Equitable Benefit-Sharing – A Fact-Finding and Scoping Study Undertaken for the Secretariat of the Convention on Biological Diversity
Comments on the draft fact-finding and scoping study		
Page #	**Para #**	**Comment**
11	17-18	**"Although the science is moving away from physical material, its use is still necessary and important for most research projects." We fundamentally disagree with the first half of the sentence. We see no evidence that science is "moving away from physical material". Science and datasets are massively expanding but this expansion does not implicitly mean that the value or importance of physical resources has coincidentally decreased.** **For example, a genomic comparison of gene content across all known bacterial and archaeal (prokaryotic) phyla (total 118) reveals that the four phyla for which we have deep representation (i.e., many species representatives), an average of 21% of the genes in any given genome are categorized as "unknown function" or "hypothetical". In comparison, the 29 phyla for which very few representatives have been obtained, the aveage unknown fraction is 42%. Finally, for the 85 phyla where no culture exists, the average unknown fraction is 64%. This unknown fraction is scientifically "useless" until representatives from these uncultured organisms are obtained and can be further phenotypically analysed. In this example, the sequence information helps scientists to know what to target next and to generate hypotheses, but it does not at all replace or diminish the importance of having the biological representatives.** ***Overmann, Abt, Sikorski (2017) Annu Rev Microbiol 71: 711*** ***Sikorski & Overmann (2017) BIOspektrum (Nov.,p. 842-843)*** **Another example comes from the pharmaceutical industry, where recent industry studies have shown that 40% of the income is currently generated by biologically-produced products. This is in drastic contrast to the early 1980s where no biologically-originated products were income-generating.** *Source: https://www.pharmasalmanac.com/articles/biologics-driving-force-in-pharma* **The main report (p.35, line 6-10) actually shows how critical physical resources are, but this importance actually seems to be diminished in the executive summary.**
2-3	All	Page numbers missing from table of contents
7 & 18	22 & 24	Discrepancy between study length – four or three months?
9	29	Suggest replacing "from" with "by"
9	29	Suggest adding "public institutes" (state, federal, or internationally government-funded institutions, but not government themselves) and adding a comma between "research institutions" and "collections". These are often separate entities, i.e., most research institutions are not collections and, conversely, many collections are not research institutions. But both do routinely produce and use DSI.
9	1-26	It is perhaps worth noting that synthetic biology is done almost exclusively with microbial genetic resources at present. And, as pointed out elsewhere in the study, but, which could be reiterated here, microbes are cosmopolitan and are widespread throughout the globe, which means that synthetic biology parts are highly unlikely to be unique or endemic to a specific country.

10	21-24	Although metadata, including geographic origin, is an important goal of the INSDC databases listed here, many have privacy policies that specifically prohibit the personal traceability of sequence because of privacy concerns. And without personal information the concept of traceability of a sequence, even if it had geographical origin information, would be impossible. Legally, it would be very difficult for the provider country to determine or prove that the sequence is truly theirs if the chain of custody (including the original sequence depositor) cannot be established.
10	35-37	There seems to be a false dichotomy here. The paragraphs seems to suggest that there are static sequence databases and then collections that separately sequence and analyse physical samples. In reality, these field collections and ex situ collections are major generators of sequence data that is quickly deposited in these same databases. Collections and citizen science programs are NOT separate actors in the system. They also contribute, use, generate, and deposit sequences in the databases. And, similarly, many university (academic researchers) access and analyse physical samples. Suggest deleting or strongly re-phrasing.
10	**38-39**	**Citation is needed for statement "Field collections of physical samples are a much smaller part of research strategies in high tech industries than they were twenty years ago." Our professional experience does not support this statement. And industry survey or study would be helpful here to support this, otherwise suggest deleting or re-wording.**
10 & 11	**38 &1-2**	**"Today, few companies undertake regular and systematic collections, although there are exceptions." This sentence, especially in context of the preceding sentences, seems to suggest that collection has decreased because physical biological samples are less important or relevant than they once were. In our experience, collections have gone down (especially in industry) NOT because the physical samples are irrelevant or have become unimportant, but rather, because of the CBD and NP, there is insufficient legal certainty and often significant bureaucratic overhead to obtain samples. Therefore, they have turned to existing collections of "safe resources" or sampling in free access countries, but have not necessarily stopped collecting.**
11	26-32	It would increase transparency, to note that the synthesis of biological molecules is not as advanced or successful as previously predicted and hoped. This is mentioned in the full report (p.36, 31-34) but not in the summary. The summary seems to imply it is commonplace
12	1-27	If BLAST requires a comparison with all sequences in a database, then how would these tools work? Would all sequences need to be acknowledged if they were used for a blast search or came up as a blast hit? These tools are interesting, but do not address the problem of using the database as a whole entity. They are only useful when an individual sequence is already identified, but this is the exception rather than rule of sequence use. **This leads to a more general question not addressed in the study (but relevant): When would "utilization" of a sequence begin? If BLAST utilization? Is phylogenetic assignement utilization? Etc…**
12	37-38	Morphological identification of all microbes is impossible. It can ONLY be done officially (taxonomic description) with accompanying sequence information of, at the very least, a full-length 16S rRNA gene.
14	34	Typo: "laboratoriews" should be "laboratories"
15	14-15	A few seems practically impossible. Sequences that are already there are free and must stay free. But then new sequences would be "siloed" and couldn't be compared or integrated. It is more than a "concern", this is a practical impossibility.
7-17		A 10-page executive summary seems very long for a 45-page report!

54 and 56	6-9 10 ff	"Additionally, homologous, or identical, sequences vital to life, and in which natural selection has eliminated mutations, might be found in different organisms around the world. This means that if companies cannot acquire legal certainty for a sequence of interest in one country, they can search for, and often find, the sequence in another country." In addition to this very valid point, it would be very difficult to assign a sequence to a certain country. Animals and plants do not observe country borders – this is particularly apparent in invasive or migratory species. Therefore a sequence obtained by sampling in a particular country could always be challenged as belonging to a different country. This would create huge uncertainty for researchers. If a researcher publishes sequence data obtained from one country in compliance with the relevant ABS regulations, another country could always challenge this and claim ownership of the sequence data. This would lead to a multitude of legal and bureaucratic issues, and even having complied with the pertaining regulations, a researcher could never be sure they would not be sued by another country.
54	38ff	"Genetic material from diverse organisms, from around the world, is commonly combined in the development of new products, processes and technologies." This is not only true for products, processes and technologies. Much of genetic research is fundamentally dependent on the use of genetic sequences from a variety of sources. This is true for basic research (creating phylogenies) as well as applied research (e.g. disease research). Forensics, e.g. in order to combat illegal trade, is also only possible based on the availability of sequences for comparison with samples confiscated by the authorities. Drawing on sequence databases to construct phylogenetic trees or trace the origin of a sample, it would simply not be possible to adhere to the ABS regulations of dozens of countries in order to cover all the sequences used.
57 and 59	27 ff 34-35	"What percentage similarity of a gene sequence requires you to consider benefit sharing? Small introduced changes can have massive effects on the genes being used, turning them from unusable to very valuable. How would this be accounted for?" This is a very important point, and one that not only applies to modifications introduced to sequences, but also to natural sequence variation. How different does a sequence have to be in order to be recognized as not belonging to a certain country? Biologists are struggling with defining thresholds between species or subspecies, because different clades differ by orders of magnitude in within- and between-taxon genetic diversity. For example, if a sequence differs by just one point mutation from another, can a country still claim it? Importantly, in this situation, the sequence could potentially stem from a different country, but the one difference could also simply represent a sequencing error. Following from this, how would one determine who can claim ownership of the "original" sequence – where would the line be drawn, and how far back would one go?

58	21 ff	Disease monitoring and research is a good example to illustrate that a) sequences cannot reliably be allocated to a certain country (globalisation means that diseases are very quickly carried around the globe), and b) restricting access to sequence information would severely hinder the development of measures to control a disease. In order to combat e.g. an ebola outbreak, it is vital to be able to trace the origins and transmission of the virus. Countries restricting access to their sequence information would in such cases incur severe negative consequences, as disease dynamics in these countries would be excluded from international studies. Since disease control plans hinge on the availability of such information, this would have disastrous effects.
62	14 ff	The question is not only distinguishing between academic and commercial research, as there are a number of additional important applications of sequence data in conservation. The use of sequences stemming from multiple sources is essential for many captive breeding programmes and – in order to have samples for comparison – for forensic analyses necessary in combatting illegal trade of plants and animals.

Please submit your comments to secretariat@cbd.int or by fax at +1 514 288 6588.