



Primary biodiversity data and the Post-2020 Global Biodiversity Framework

Andrew Rodrigues, David Bloom, Paula Zermoglio, Robert Guralnick, Tim Hirsch,
Jillian Campbell, Natasha Ali, Simon Ferrier, Aidin Niamir, Maria Cecilia Londoño,
Yanina Sica

Table of Contents

Preface	1
1. Data models and standards for improved usability	1
2. Spatial, temporal and taxonomic biases in the data	2
3. Transparency of data and methodologies used in indicators	3
References	4
Colophon	6
Suggested citation	6
Contributors	6
Licence	6
Persistent URI	6
Document control	6
Cover image	7

Preface

For decades scientists and policymakers have called for more effective and efficient methods to monitor and address the global challenges of biodiversity loss and environmental degradation (Carpenter et al. 2006, Pereira & Cooper 2006, Navarro et al. 2017). The Convention on Biological Diversity (CBD) is developing a post-2020 global biodiversity framework to support an accelerated push for action with a new set of goals and targets that will replace the Aichi targets. The First Draft of the framework (CBD 2021a) built upon the mission of the Strategic Plan for Biodiversity 2011-2020 (CBD 2010), retains the 2050 Vision and proposed four long-term goals with associated outcomes for 2030 and 2050, including 21 action targets through which the goals should be achieved.

As part of the post-2020 framework, Parties to the CBD will agree on a subset of headline quantitative indicators that will allow Parties to monitor progress towards the goals and action targets (UNEP-WCMC & BIP 2020). These indicators should apply to all countries and allow prioritization of capacity and resource needs. A sufficient supply of high-quality data and a range of Essential Biodiversity Variables (EBVs), a framework that defines a minimum set of critical variables required to study, report and manage biodiversity change (Jetz et al. 2019), would need to be in place to ensure the comparability of headline indicators across the globe and allow scalable reporting across space and time (CBD 2021b). These need to be supported by common methodologies and data standards, and biodiversity observation networks and information facilities.

GBIF commissioned a study in 2020 to analyse those indicators (synthetic or derived metrics) available at that time that make use of primary biodiversity data (raw observations) to support the post-2020 biodiversity framework. The study reviewed and characterized the sources of primary biodiversity data, identifying where data use is redundant, and how GBIF might mobilize data more effectively to support the implementation of the new post-2020 framework. To ensure a robust monitoring framework three key elements need to be addressed: the fitness of data models and standards for the development of indicators; biases within the data that could prevent the utility of indicators at different reporting scales; and a lack of transparency in the way in which data is applied in indicators, as well as in the provenance of the data generating indicator results. The main conclusions of the study are set out in this paper and will be addressed in an online consultation.

1. Data models and standards for improved usability

Over the last two decades there have been enormous efforts to mobilize biodiversity data, which have resulted in the availability of massive amounts of published data that can be readily discovered, accessed and freely used for onward applications. The process of building meaningful EBVs that can inform indicators needs data that can be reliably tracked across not just organism, space and time but also provenance; the latter includes relevant, complete and searchable metadata about the inventory process and the methods that produced those data. Much of the data shared through biodiversity data platforms lack one or more of those four components, which limits or excludes their use in the creation of EBVs and biodiversity indicators. Furthermore, much of the data currently shared correspond to incidental records and lack any defined inventory or survey methods.

For EBVs that use multi-variable analyses to aggregate and homogenize data across species, space

and time, a taxon name, an event date and a set of coordinates are not enough to account for any bias or deficiencies in the available data. One way to help to overcome these biases is to publish occurrence and event records with metadata that describes the collection methodology and processes that are as rich as possible. However, this type of correction will only be useful for certain types of analyses. Many species occurrence records that represent only the presence of the species (i.e. incidental records) will still not be useful for EBVs that must account for data that enables inference about absence of species. For these EBVs, well-documented monitoring or inventory event data is needed.

As more monitoring data becomes available, expanded best practice guidelines should include, but may not be limited to, how to share quality metadata containing details of the sampling methods employed, the scope, and descriptions and provenance of the collected data. To make this practical, biodiversity data platforms will need to review and amend current data-sharing standards and practices, and upgrade their infrastructures to host and display new types of data and data formats such as is the case with [GBIF's ongoing consultation to review its current data model](#). An example of a new standard that is under review for implementation is the Humboldt extension to Darwin Core ([Guralnick et al. 2017](#), [Sica & Zermoglio 2021](#)). Furthermore, data publishing institutions could be encouraged to create "sub-collections" of their data that meet these metadata requirements that they could publish separately from their larger corpus of data. An increased focus on the publication of past and current monitoring and inventory datasets with the expressed purpose of supporting EBV and biodiversity indicator creation would require strengthened ties with the research and monitoring communities that produce those data.

2. Spatial, temporal and taxonomic biases in the data

According to the [UNEP-WCMC & BIP 2020](#), "The post-2020 global biodiversity framework will be implemented primarily at the national level. It is therefore important that the relative roles and suitability of both global and national indicators are considered." The provision of data suitable for the national-level implementation strategy of the post-2020 framework that addresses the challenges of scalability will require biodiversity data platforms to improve the quality and completeness of available data. As noted in a recommendation of the 3rd meeting of the CBD's Subsidiary Body on Implementation (SBI3), addressing knowledge management for the GBF, this will involve the establishment of biodiversity observation networks and information facilities, supported by data-sharing policies, associated capacity-building and guidance, to underpin the generation of the information needed to implement and track the goals and targets of the global biodiversity framework ([CBD 2022](#)).

Bias refers to a systematic lack of information due to a sampling design that relies on incorrect assumptions, which may be taxonomic, geographic, temporal or environmental. One cause of bias is the lack of capacity for making existing data accessible from particular regions or across taxonomic groups. A number of initiatives have taken steps to address these biases in global datasets. For example, GBIF's Biodiversity Information for Development (BID) and Biodiversity Information Fund for Asia (BIFA) programmes, co-funded respectively by the European Union and the Ministry of Environment, Japan, have made significant efforts to increase capacity for mobilizing data from institutions in Africa, the Caribbean, the Pacific and Asia, and to fill data gaps in those regions. Recent

guidelines on the publication of DNA-derived data through GBIF allow for the integration of data from environmental DNA sampling, and help to increase data coverage in data-poor ecosystems and taxonomic groups. And innovative uses of the data, such as the Bioclimatic Ecosystem Resilience Index (BERI) (Ferrier et al. 2020), are able to assess changes in biodiversity over time without a full time-series of observations and thus respond to temporal biases within the data.

There are also opportunities to mobilize more data, especially monitoring and inventory data at the local level. Historically, much of the biodiversity science community has been focused on the mobilization of data within established legacy collections, such as those in museums, laboratories and government agencies (Guralnick et al. 2007). There has now been a shift toward monitoring and observation projects, including citizen science. As pressure mounts to address questions about the status and trends of biodiversity at different scales, it is these data from local sources focused on the smaller-scale monitoring of national parks, waterways, and wildlands - data often collected by indigenous peoples and local communities with local knowledge - that are of critical importance in efforts to fill knowledge gaps and maintain on-going monitoring (Tengö et al. 2017, Hill et al. 2020, Brook & McLachlan 2008, Geldmann et al. 2021).

The private sector is also an important source of biodiversity data in the form of environmental assessments, impact assessments, and other project-based analyses. Increasing numbers of private sector actors are publishing biodiversity data through GBIF and the GBIF community is engaging with the private sector directly through several initiatives, such as Data4Nature which targets public development banks to encourage data sharing as part of financing conditions. In addition, some national governments have begun to mandate private sector data publication, and financial institutions have created incentives for commercial entities to share non-sensitive data with GBIF and other national and global repositories (Equator Principles Association 2020).

3. Transparency of data and methodologies used in indicators

Assessing the quality of primary biodiversity data that meets the standards needed for further use in indicators is critical. Even if we were to satisfy the need for better quality data, questions would remain about how homogeneous and repeatable the treatment of the same data can be in different contexts. The same data, from multiple sources, is being used by distinct organizations or collaborations to build EBVs and indicators. Stakeholders developing a given EBV or indicator treat the data independently, apply their own filters and quality checks, and perform their own taxonomic harmonization process, which may be more or less similar to those used by other stakeholders. If biodiversity data platforms could prepare and share species occurrence data in advance for EBV and indicator creation, as EBV-usable datasets, or make the workflows to process data available for example, better consistency and transparency might be achieved. GBIF is exploring ways of assisting this process, for example through pre-filtered versions of GBIF-mediated data exported regularly to public cloud environments.

A second opportunity of equal importance is to improve the communication pipeline between data provider and data user. Data and communications about these data, tend to flow in one direction, from local data collection and mobilization to scientists and policymakers, with little to no communication in the opposite direction. GBIF and other biodiversity data platforms have made commendable efforts to track downloads of data and to report the citations of published works back

to data publishers when they are made public through the use of Digital Object Identifiers (DOIs). Improved communications build trust across the data provider network by communicating back to organizations and individuals at the local level about the uses of data. These communications could occur in many ways, including notifications that alert data publishers when their data have been used in the creation of EBVs, biodiversity indicators and other high-level policy documents, using tools similar to the GBIF citation widget. Another effective communication strategy could be the presentation of specific examples that demonstrate how high-quality data and associated metadata are being used to influence science and policy as a part of capacity-building activities and other public events. These possibilities will remain only possibilities, however, without greater transparency.

A third opportunity to work towards greater transparency and traceability across the entire information supply chain is to document all steps taken to create indicators. In this complex process, it is not uncommon for the processes and analyses used to generate these synthesized data and policy products to remain undocumented or hidden from public view. Similarly it is equally difficult to know exactly which data were used in the processes and how. The CBD Secretariat and UNEP-WCMC, are currently working on standardizing the metadata requirements for the proposed headline indicators (see example for the Species Habitat Index, [UNEP-WCMC 2021](#)); this must include clear reporting of datasets (DOIs) used and data providers consulted to improve traceability even further.

References

- Brook RK & McLachlan SM (2008) Trends and prospects for local knowledge in ecological and conservation research and monitoring. *Biodiversity and Conservation* 17: 3501-3512. <https://doi.org/10.1007/s10531-008-9445-x>
- Carpenter SR, DeFries R, Dietz T et al. (2006) Millennium Ecosystem Assessment: Research Needs. *Science* 314: 257-258. <https://doi.org/10.1126/science.1131946>
- Equator Principles Association (2020) The Equator Principles. July 2020. Available at: <https://equator-principles.com/wp-content/uploads/2021/02/The-Equator-Principles-July-2020.pdf>
- Ferrier S, Harwood TD, Ware C & Hoskins AJ (2020) A globally applicable indicator of the capacity of terrestrial ecosystems to retain biological diversity under climate change: the bioclimatic ecosystem resilience index. *Ecological Indicators* 117: 106554. <https://doi.org/10.1016/j.ecolind.2020.106554>
- Geldmann J, Deguignet M, Balmford A et al. (2021) Essential indicators for measuring site-based conservation effectiveness in the post-2020 global biodiversity framework. *Conservation Letters* e12792. <https://doi.org/10.1111/conl.12792>
- Guralnick RP, Hill AW & Lane M (2007) Towards a collaborative, global infrastructure for biodiversity assessment. *Ecology Letters* 10(8): 663-672. <https://doi.org/10.1111/j.1461-0248.2007.01063.x>
- Guralnick R, Walls R & Jetz W (2017) Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. *Ecography* 40: 001-012. <https://doi.org/10.1111/ecog.02942>
- Hill R, Adem C, Alangui WV et al. (2020) Working with Indigenous, local and scientific knowledge

in assessments of nature and nature's linkages with people. *Current Opinion in Environmental Sustainability* 43: 8-20. <https://doi.org/10.1016/j.cosust.2019.12.006>

- Jetz W, McGeoch MA, Guralnick R et al. (2019) Essential biodiversity variables for mapping and monitoring species populations. *Nature Ecology and Evolution* 3: 539–551. <https://doi.org/10.1038/s41559-019-0826-1>
- Navarro LM, Fernández N, Guerra C et al. (2017) Monitoring biodiversity change through effective global coordination. *Current Opinion in Environmental Sustainability* 29:158–169. <https://doi.org/10.1016/j.cosust.2018.02.005>
- Pereira HM & Cooper HD (2006) Towards the global monitoring of biodiversity change. *Trends in Ecology and Evolution* 21(3): 123-129. <https://doi.org/10.1016/j.tree.2005.10.015>
- Secretariat of the Convention on Biological Diversity (2010) Strategic Plan for Biodiversity 2011-2020, including Aichi Biodiversity Targets. <https://www.cbd.int/sp/>
- Secretariat of the Convention on Biological Diversity (2021a) First Draft of the Post-2020 Global Biodiversity framework. <https://www.cbd.int/doc/c/914a/eca3/24ad42235033f031badf61b1/wg2020-03-03-en.pdf>
- Secretariat of the Convention on Biological Diversity (2021b) Report of the virtual sessions of the fifth science-policy forum for biodiversity and the eighth international conference on sustainability science. <https://www.cbd.int/doc/c/8cd2/6eab/663d8a4cc2d198b104225345/sbstta-24-inf-28-en.pdf>
- Secretariat of the Convention on Biological Diversity (2022) Knowledge Management and the Clearing-House Mechanism. <https://www.cbd.int/doc/c/bc3c/fa25/e5d7928206fe7f3fc20aa9b1/sbi-03-l-21-en.pdf>
- Sica YV, Zermoglio PF (2021) Unlocking Inventory Data Capture, Sharing and Reuse: The Humboldt Extension to Darwin Core. *Biodiversity Information Science and Standards* 5: e74275. <https://doi.org/10.3897/biss.5.74275>
- Tengö M, Hill R, Malmer P et al. (2017) Weaving knowledge systems in IPBES, CBD and beyond—lessons learned for sustainability. *Current Opinion in Environmental Sustainability* 26-27: 17-25. <https://doi.org/10.1016/j.cosust.2016.12.005>
- UNEP-WCMC & BIP (2020) Indicators for the post-2020 global biodiversity framework. <https://www.cbd.int/sbstta/sbstta-24/post2020-indicators-en.pdf>
- UNEP-WCMC & BIP (2021) Indicator metadata sheet - Species Habitat Index. https://cdn.mol.org/static/files/indicators/habitat/WCMC-species_habitat_index-15Feb2022.pdf

Colophon

Suggested citation

Bloom D, Zermoglio P, Guralnick R, Rodrigues A, Hirsch T, Campbell J, Ali N, Ferrier S, Niamir A, Londoño MC & Sica Y (2022) Primary biodiversity data and the Post-2020 Global Biodiversity Framework. GBIF Secretariat: Copenhagen. <https://doi.org/10.15468/doc-pgg2-xn60>

Contributors

- David Bloom, VertNet: UF/Florida Museum of Natural History
- Paula Zermoglio VertNet
- Robert Guralnick, Florida Museum of Natural History
- Andrew Rodrigues, GBIF Secretariat
- Tim Hirsch, GBIF Secretariat
- Simon Ferrier, CSIRO Land and Water
- Aidin Niamir, Senckenberg Biodiversity and Climate Research Centre
- Yanina Sica, Yale University
- Natasha Ali, United Nations Environment Programme World Conservation Monitoring Centre (UNEP-WCMC)
- Jillian Campbell, UN Convention on Biological Diversity
- Maria Cecilia Londoño, Humboldt Institute | Group on Earth Observations Biodiversity Observation Networks (GEO BON)

Additional contributors to subsequent versions will be credited here.

Licence

The document *Primary biodiversity data and the Post-2020 Global Biodiversity Framework* is licensed under [Creative Commons Attribution-ShareAlike 4.0 Unported License](#).

Persistent URI

<https://doi.org/10.15468/doc-pgg2-xn60>

Document control

Adapted from Bloom DA, Zermoglio P & Guralnick R (2021) [Analysis of biodiversity data needs in the post-2020 framework](#). Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-2ph8-0403>.

Cover image

Bleeding bonnet (*Mycena sanguinolenta*), observed in Norway. Photo © 2020 Kirsti Anne Mandal via [Norwegian Species Observation Service](https://www.gbif.org/occurrence/2848866491).