



# Analysis of biodiversity data needs in the post-2020 framework

David A Bloom, Paula Zermoglio & Robert Guralnick

Version 1a247a5, 2022-02-14 08:56:57 UTC

# Table of Contents

Colophon .....	1
Suggested citation .....	1
Authors .....	1
Licence .....	1
Persistent URI .....	1
Document control .....	1
Cover image .....	1
1. Background and Introduction .....	1
2. Methods .....	5
3. Results .....	6
4. Discussion .....	15
4.1. Relevant and timely high-quality data .....	16
4.2. Support the network .....	18
5. Conclusions .....	21
Appendix A: Example .....	23
References .....	24

# Colophon

## Suggested citation

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>

## Authors

David A Bloom, Paula Zermoglio & Robert Guralnick

## Licence

The document *Analysis of biodiversity data needs in the post-2020 framework* is licensed under [Creative Commons Attribution-ShareAlike 4.0 Unported License](#).

## Persistent URI

<https://doi.org/10.35035/doc-2ph8-0403>

## Document control

First edition, September 2021

## Cover image

Ensatina (*Ensatina eschscholtzii*), Sonoma County, California, United States. Photo 2021 Elliott Smeds via [iNaturalist research-grade observations](#), licensed under [CC BY-NC 4.0](#).

# 1. Background and Introduction

For decades scientists and policymakers have been calling 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) has been a primary force behind the creation of systems to monitor changes in biodiversity and has served as a global assessor of the progress made toward the achievement of goals, such as the [Aichi Biodiversity Targets established in 2010](#). Despite wide-spread efforts to stem the tide of ecological decline, the CBD declared in its Global Biodiversity Outlook 5 (CBD 2020c) that none of the 20 Aichi targets for biodiversity were fully achieved by the 2020 deadline. Only six of these targets were accomplished partially.

The failure of the global community to rise to the call of the CBD has been attributed to many factors, including gaps in and limited access to scientific knowledge, regional economic disparity, and a lack of political will (CBD 2020c, CBD 2020e, IPBES 2019). Widespread use of sensor-based

observation systems, the rise of global biodiversity aggregation services, and novel applications for genetic and genomic discovery have helped to address some of these challenges (Turner 2014, Bush et al. 2017). Primary ecological and biodiversity-based data are far more discoverable than in years past (Constable et al. 2010). The increase in data availability is necessary, but insufficient to meet the targets set to sustain natural capital.

To address the pressing crisis in biodiversity the CBD is developing a post-2020 global biodiversity framework to support an accelerated push for action with a new set of goals and targets to replace the Aichi targets. The Zero Draft of the framework (CBD 2020a, CBD 2020b) builds upon the mission of the Strategic Plan for Biodiversity 2011-2020 (CBD 2010), retains the 2050 Vision and proposes five long-term goals with associated outcomes for 2030 and 2050, including 20 action targets through which the goals should be achieved.

Access to reliable, high-quality information is essential to achieve the goals of the post-2020 framework. The new Target 19 proposes to expand upon the partial successes towards achieving its previous version to “ensure that quality information, including traditional knowledge, is available to decision makers and public [sic] for the effective management of biodiversity through promoting awareness, education and research.” This information is indispensable, as the proposed framework presupposes that progress to attain the goals will be monitored over time. The assessment, development and use of appropriate biodiversity indicators (UNEP-WCMC & BIP 2020) will play an important role in the reporting processes.

A biodiversity indicator, as defined by the Biodiversity Indicators Partnership (BIP 2011), is “a measure based on verifiable data that conveys information about more than itself”. Indicators provide a means to monitor and report on the state of biodiversity, the Earth’s ecological systems and the status of efforts to protect and preserve them. The BIP develops and ratifies key biodiversity indicators (BIP 2011; UNEP-WCMC & BIP 2020) with the explicit purpose to assess the progress and success of the draft monitoring framework (CBD 2020d) of the post-2020 global biodiversity framework (CBD 2020d). Indicators inform policymakers to make justified decisions, measure the progress of national and regional efforts, raise awareness, and promote educational efforts (CBD 2006). Some indicators have been used for years and applied across different biodiversity-related conventions and processes (e.g. SDGs, IPBES, CITES, etc., see Moul et al. 2018), while others are derived from the combined assessment of other indicators and indices.

- The Zero Draft of the framework (CBD 2020a, CBD 2020b) builds upon the mission of the Strategic Plan for Biodiversity 2011-2020 (CBD 2010), retains the 2050 Vision and proposes five long-term goals with associated outcomes for 2030 and 2050, including 20 action targets through which the goals should be achieved.
- Access to reliable, high-quality information is essential to achieve the goals of the post-2020 framework. The new Target 19 proposes to expand upon the partial successes towards achieving its previous version to “ensure that quality information, including traditional knowledge, is available to decision makers and public [sic] for the effective management of biodiversity through promoting awareness, education and research.” This information is indispensable, as the proposed framework presupposes that progress to attain the goals will be monitored over time. The assessment, development and use of appropriate biodiversity indicators (UNEP-WCMC & BIP 2020) will play an important role in the reporting processes.

In the context of the post-2020 framework a subset of headline quantitative indicators should be

prioritized (UNEP-WCMC & BIP 2020). Such indicators should apply to all countries and allow prioritization of capacity and resource needs. A sufficient supply of high-quality data, common methodologies and data standards, and observation systems, must be available to ensure that headline indicators are comparable across the globe and to allow scalable reporting across space and time (CBD 2021).

Timely publication of high-quality, comprehensive data is pivotal for indicators to be broadly applicable in aiding decision-making (Collen & Nicholson 2014, Stevenson et al. 2021). Data may suffer from a lack of completeness and from bias. A lack of completeness may result from two different aspects of the data acquisition and sharing workflow. First, data components may not be available and/or shared (e.g. missing fields in a database). Second, completeness may refer to geographic or temporal gaps during sampling events (e.g. a survey to determine presence of a given set of taxa at a national level where certain regions in the country are partially or not represented). Bias here 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. Overall, providing timely high-quality data should take into consideration ways to address both data completeness and data bias (Nicholson et al. 2012). Furthermore, to ensure timeliness, data should conform to FAIR principles (Findability, Accessibility, Interoperability, and Reuse of digital assets, Wilkinson et al. 2016) and encourage discoverability and accessibility to the information.

Biodiversity and ecological data are necessary for the creation of CBD's biodiversity indicators to monitor progress to protect and preserve the Earth's natural systems. The primary biodiversity data themselves, however, are not fit for use by policymakers and must first pass through the early stages of an information supply chain (Kissling et al. 2018a) to become harmonized into derived data products that are ready for consumption. At the heart of the supply chain, between the primary biodiversity data and the indicators, sit a diverse set of analyses and processes that convert data from a wide-range of heterogeneous sources into homogenous and harmonized data products. These data products may be used in the generation of Essential Biodiversity Variables (EBVs), which, in turn, may be used to support the creation of biodiversity indicators. As the name implies, EBVs are an essential component of the information supply chain, yet their quality and usefulness are highly dependent upon the quality and quantity of the biodiversity data available for use (Kissling et al. 2018a, Kissling et al. 2018b, Navarro et al. 2017, Proença et al. 2017).

Essential Biodiversity Variables (EBVs) are measured and used to evaluate changes in the biological state across space, time and biological organization (Schmeller et al. 2017, Hardisty et al. 2019). The measurement of some of these variables may result from the integration of multiple data sources, such as species occurrence data and remote-sensing observations. These data sources are aggregated and then assessed to evaluate a variable that describes temporal and spatial change that could not be derived from a single source alone (REF?). Fundamental characteristics of EBVs are the measurements that allow comparisons and assessment of change across various scales of time and space. In this context, monitoring initiatives such as the Earth Observation Networks (EONs) and Biodiversity Observation Networks (BONs) play a pivotal role in the acquisition of data and delivery of products that can be readily used in policy- and decision-making (Lindenmayer et al. 2017, Scholes et al. 2008, GEO BON 2021). Despite ongoing efforts, challenges remain regarding the extent of species monitoring programs across different taxa and regions and their integration with the enormous amount of incidental occurrence data available through global biodiversity data platforms, such as the Global Biodiversity Information Facility (GBIF) (Pereira et al. 2010, Navarro et al. 2017).



Scientists continue to struggle to find sufficient quantities of occurrences that contain the high-quality taxonomic and geographic data necessary to create EBVs and biodiversity indicators (Ariño et al. 2013, Proença et al. 2017), despite the abundance of occurrence data (e.g. GBIF currently provides access to around 1.7 billion occurrence records). Leaving aside considerations of the quality of the available data, occurrences for many geographic areas or species are simply lacking, due, for example, to a lack of capacity to mobilize the data, limited access to difficult terrain, or restricted monitoring or inventory efforts (Sousa-Baena et al. 2014, Feeley 2015). Further, relevant details are often missing from the data, or they are captured in a way that renders them challenging to find, parse and process. These missing details are frequently, but not exclusively, related to associated sampling methods and taxonomic and geographic coverages. This information is essential to account for sources of bias and uncertainty in the data used to generate an EBV. Data and metadata gaps may result from historical data collection practices, collecting for purposes other than monitoring, loss of data over time or the exclusion of data during the publication process (e.g. the flattening of metadata to fit existing data-sharing standards and infrastructures, Peterson et al. 2018).

GBIF is an international network and data infrastructure funded by the world's governments to provide anyone, anywhere with open access to data about all types of life on Earth. Species occurrence data aggregated by GBIF have been used in research, education, and policy-making since it was established in 2001. These data have been used in nearly 6,000 peer-reviewed publications since 2008 (GBIF Secretariat 2021). This primary biodiversity data also plays a critical role in the generation of information resources that guide the decision-making processes and the monitoring of progress to meet international commitments under the CBD and other multilateral environmental agreements. Growth in species occurrence records accessible through GBIF data serves as a primary indicator for tracking progress towards Target 19 of the Aichi Biodiversity Targets. GBIF-mediated data also contribute to the development of several other indicators covering progress towards Aichi targets, including Target 5 (habitat loss and degradation), Target 9 (invasive alien species), Target 11 (protected areas), Target 12 (extinction risk) and Target 13 (genetic diversity).

The process to develop and approve the post-2020 global biodiversity framework presents GBIF with a unique opportunity, especially given concerns that have been raised since the publication of the zero-draft and its update (CBD 2020a, CBD 2020b). In particular, some stakeholders have called for revisions to certain goals, targets and indicators to ensure the framework's effectiveness in different thematic areas and its overall goals (see Díaz et al. 2020, Essl et al. 2020, Hoban et al. 2020, Williams et al. 2020). As the premier data-sharing infrastructure for global biodiversity data, GBIF may position itself both to influence and to contribute to new and robust indicators of biodiversity while continuing to inform existing indicators. Biodiversity-related indicators and EBVs — and the ability to use them to achieve goals and targets — will benefit from high-quality biodiversity data brought together through GBIF.

When COP15 delegates approve the post-2020 framework, a new wave of biodiversity-related activities will begin; existing biodiversity indicators will be refreshed, new indicators may be constructed. To inform these indicators, primary biodiversity data and a range of EBVs and methods of analyses will be needed. GBIF, as an observer of the CBD, has been part of this conversation and has taken steps to ensure that GBIF, the members of the *alliance for biodiversity knowledge* and all of the biodiversity data this collaboration has mobilized are prepared to contribute to the implementation of the post-2020 framework in meaningful ways. Key to these efforts is a better understanding of the landscape of data sources and relevant stakeholders involved in the

development of indicators, as well as the dependencies among them, to help direct future collaborations and commitments.

In this study we identify projects and products that make use of primary biodiversity data to support the post-2020 biodiversity framework and how they inform indicators and information tools that address different CBD concerns. Also, we review and characterize the sources of primary biodiversity data used to inform indicators and other information products to identify where data use is redundant and how GBIF might provide data more effectively. Finally, we perform an analysis of the likely dependencies on primary biodiversity data within the post-2020 biodiversity framework, including primary biodiversity data and data from other disciplines, with an assessment of the intervening organizations and their roles in data collection, harmonization and delivery of primary biodiversity data, EBVs and indicators towards policy agendas.

## 2. Methods

To ensure efficient and effective provision of biodiversity data and services, GBIF commissioned an analysis of existing biodiversity indicators that depend on primary biodiversity data. This analysis looks at how GBIF-mediated data feeds into the development of indicators and how it informs the work and products of the **IPBES**, the CBD, and other global, national and regional science-policy processes.

A data source is defined broadly in this study. Most data sources are products generated for use by researchers, policymakers and others across multiple communities. They are often used as baseline or primary data inputs for analyses. Once these analyses are performed the result is a data product that contains modified data from the initial data sources, as well as new data derived from the analyses. Some EBVs and indicators use only the primary biodiversity data from various data sources, while others may use a combination of primary data and data products in their creation. Thus, the term “data sources” refers to both primary data and data products used to create indicators. For the purposes of this study, data aggregated through GBIF is considered to be primary data, and not as a data product.

This study began with an exploration of the existing indicators identified as most likely to contribute to the monitoring framework of the post-2020 global biodiversity framework. A first step was to modify the list of **BIP** and **GEO-BON** indicators published in the Information Document prepared for SBSTTA24 by UNEP-WCMC and BIP (**UNEP-WCMC & BIP 2020**). This list contained indicators predicted to contribute to the post-2020 framework directly. Other indicators identified as potential contributors were added to generate a comprehensive list. We selected 11 indicators from the extensive list for in-depth assessments of their use of primary biodiversity data. These indicators were selected for analysis because they declared the use of GBIF-mediated data or partnerships with GBIF in their online documentation, such as the indicator descriptions on the BIP web site. Two additional indicators not on the SBSTTA24 list were found to have direct connections to GBIF and were added to the analysis (Growth of Species Occurrence Records Available through GBIF and the Species Status Information Index). Other indicators, such as the IUCN's Red and Green Lists may also use GBIF data, but no evidence of use was available publicly, so those indicators were not included in this study. The list of indicators assessed can be found in **Table 1**.

The Information Document prepared for SBSTTA24 (**UNEP-WCMC & BIP 2020**) provided a limited quantity of information about each of the selected indicators. Documentation describing each

indicator provided on the BIP web site was also limited. Therefore, we undertook an investigation to better understand

1. the specific characteristics of the data sources used in each indicator
2. the types of data that contributed to the indicators (in addition to biodiversity data)
3. the actual datasets used to build each indicator

For each source we recorded a broad range of characteristics, including:

- the geographic, temporal and thematic spread of the source
- its funding sources
- the origin of the data aggregated into the source (e.g. repositories gathering data from natural history collections; citizen science data; data produced from research activities directly)
- the data provenance (e.g. indicating whether it is known, declared or not)
- the kind of access that is granted to the source (e.g. open, free, licensed, variable)
- the source activity/currency (e.g. if it is an ongoing activity or a one-time-only release)

Where available, we recorded links to the primary biodiversity datasets.

To identify the types of data that each indicator uses and the dependencies between the agencies using those data, we categorized the indicators by use of:

1. species occurrence data (records/data that describe the presence or absence of species)
2. species information (descriptions of location or traits of species, including range maps and distributions)
3. abiotic information (earth science data, non-biological data)
4. genetic (data about genetic resources)
5. other (data that don't fit into the above categories neatly).

To gather this information, we consulted the source organizations or projects through their websites and data portals. Direct communications clarified methodological approaches and integration with other data sources. We paid special attention to data accessibility and the reproducibility of the results, recording any issues that arose in defining the provenance of the data.

## 3. Results

As we started to assess the 11 biodiversity indicators, several patterns emerged almost immediately and held true for the duration of the analysis (Table 1 and Table 2). These patterns can be grouped into four themes that influence how biodiversity indicators are created and how the groups responsible for the maintenance of each indicator communicate with the broad community.

1. Transparency influences both how EBVs and indicators are created, and how they are shared with the research and policy communities. Individuals attempting to understand how any given indicator was formulated, how it made use of primary biodiversity data, and which processes and analyses were applied to those data, have few clear pathways to find answers. Thus, many EBVs



and indicators present themselves as ‘black boxes’ and require specialized knowledge to understand their inner workings. We present specific examples in the Discussion section below.

2. The data sources used to create biodiversity indicators represent a broad diversity in topic and presentation (Table 1). We observed several strong patterns in the data sources and their host organizations (Table 1 and Table 2).
  - a. Biodiversity data platforms are generally the primary sources of the biodiversity data used in indicator creation. These platforms, such as GBIF.org for biodiversity data, differ from the original publishers of the data in that they are not usually the stewards or sources of the primary biodiversity data. Organizations such as natural history museums, government collections, and non-governmental organizations often provide the primary data and publish them to biodiversity data platforms to ensure that these data have broad distribution and are easily discoverable. Platforms provide the broad community with access to open, licensed data that tend to be general in their taxonomic, geographical and temporal scopes, although there is significant variation in the quantity, quality and presentation of the data from one platform to another.
  - b. The data sources used to construct indicators are not limited to those provided as primary biodiversity data. Biodiversity data are used in combination with sources of other types of data from a variety of sources, such as distribution ranges and maps, land cover and land use imagery, climatic information, geographic administrative and political information, and socio-economic information. It is very common for biodiversity data to be used in conjunction with geographic and land use data, thus the quality of the locality data within each data set correlates with the quality of the output of the analyses performed for a given indicator.
  - c. Most primary data sources used in the indicators reviewed are freely accessible, yet the documentation of and access to the specific data actually used in the construction of the biodiversity indicators is often not available. This limitation presents challenges for the transparency of the creation processes and for communicating about each indicator (related to theme 1 above). A significant proportion of the data used in building some of the indicators comes from research published in closed-access scientific journals.
  - d. In almost all cases, the data sources used in indicator construction are generated, maintained and aggregated by governmental agencies or non-governmental organizations. Data that originate from or are maintained by private entities may be used in conjunction with public data, but there is little evidence to link any for-profit group to indicator creation.

Table 1. Summary of data sources types, organizations and data types feeding into indicators. For more details, see [Supplemental Materials](#).

<b>Indicator</b>	<b>Organization responsible for indicator</b>	<b>Identified data sources type</b>	<b>Data source organization</b>	<b>Data type</b>
Growth of Species Occurrence Records Accessible Through GBIF	GBIF	occurrence datasets	GBIF	occurrence

<b>Indicator</b>	<b>Organization responsible for indicator</b>	<b>Identified data sources type</b>	<b>Data source organization</b>	<b>Data type</b>
<b>Species Status Information Index (SSII)</b>	GEO-BON	occurrence datasets	GBIF	occurrence
	Map of Life	research outputs	Map of Life*	occurrence, species information
<b>Species Protection Index</b>	GEO-BON Map of Life	landscape datasets	Landsat/MODIS	landscape
		peer-reviewed publications, research outputs	Map of Life*	occurrence
		research outputs, occurrence datasets	GBIF	occurrence
<b>Species Habitat Index (SHIs)</b>	Map of Life, Yale University NGS	landscape datasets	Landsat/MODIS	landscape
		peer-reviewed publications, research outputs	Map of Life*	occurrence
		research outputs, occurrence datasets	GBIF	occurrence
<b>Biodiversity Habitat Index</b>	CSIRO	peer-reviewed publications	n/a	species information
		peer-reviewed publications, landscape datasets	n/a	species information, abiotic info
		occurrence datasets	GBIF	occurrences
<b>Bioclimatic Ecosystem Resilience Index (BERI)</b>	CSIRO	occurrence datasets	GBIF	occurrence
		peer-reviewed publications, landscape datasets, outputs	Map of Life*	occurrence
		abiotic datasets	WorldClim	abiotic info
		abiotic datasets	Soil Grids	abiotic info
		abiotic datasets	EarthEnv	abiotic info
		abiotic datasets	WorldGrids	abiotic info

Indicator	Organization responsible for indicator	Identified data sources type	Data source organization	Data type
Protected Area Representativeness Index	CSIRO +(GEOBON, GBIF, Map of Life)	landscape datasets	Landsat/MODIS	species info, abiotic info
		research outputs, occurrence datasets	GBIF	occurrences
		landscape datasets	Protected Planet	other
		peer-reviewed publications, landscape datasets	n/a	species info, abiotic info
		abiotic datasets	WorldClim	abiotic info
		abiotic datasets	Soil Grids	abiotic info
		abiotic datasets	EarthEnv	abiotic info
		abiotic datasets	WorldGrids	abiotic info
Crop Wild Relative Index	Alliance Bioversity CIAT & IUCN/CW RSG	occurrence datasets	GBIF**	occurrences
		landscape datasets	**	species info, abiotic info
		genetic datasets	**	genetic data

<b>Indicator</b>	<b>Organization responsible for indicator</b>	<b>Identified data sources type</b>	<b>Data source organization</b>	<b>Data type</b>
<b>Agrobiodiversity Index</b>	Alliance Biodiversity & CIAT	other	Alliance Biodiversity & CIAT	other
		other publications	ESDAC	species information
		landscape datasets	ESA	Landscape
		research outputs	n/a	species information
		peer-reviewed publications	n/a	occurrence, species information, abiotic information, landscape, other
		occurrence datasets	CIAT	species information
		occurrence datasets	GBIF	occurrence
		genetic datasets	Genesys	genetic data
		other	Alliance Biodiversity & CIAT	other
		other	OECD	other
		occurrence datasets, genetic datasets, landscape datasets, abiotic datasets, peer-reviewed publications, research outputs, other	Yale University	occurrences, species information, abiotic information, genetic data, landscape, other
genetic datasets, other	FAO	genetic data, abiotic information, other		

Indicator	Organization responsible for indicator	Identified data sources type	Data source organization	Data type
Comprehensiveness of conservation of socioeconomically as well as culturally valuable species	CIAT, Crop Trust	occurrence datasets	GBIF**	occurrence
		peer-reviewed publications, other biological datasets	World Economic Plants database	species information
		genetic datasets	GENESYS	genetic data
		occurrence datasets	Crop Wild Relatives database of Global (CWR) Project	occurrence
		abiotic datasets	WorldClim	abiotic info
		abiotic datasets	CGIAR-CSI SRTM	abiotic info
		abiotic datasets	ISO	abiotic info

Table 2. Summary of the characteristics of the data sources organizations. For more details, see Supp. Material.

Data source organization	Geographic spread	Taxonomic spread	Temporal spread	Funding origin	Data origin	Data provenance	Access	Activity
CGIAR-CSI SRTM	global	n/a	n/a	government, NGO	government, NGO, research	not declared	open	active
CIAT	global	agricultural taxa	1967-2020	government, NGO	government, research, other	not declared	open	active
Crop Wild Relatives database of Global (CWR) Project	global	plant taxa	n/a	government, NGO	government, NGO, research	not declared	open, licensed	active
EarthEnv	global	n/a	n/a	government, NGO	government, NGO, research	not declared	not specified	active
ESA	global	n/a	1992-2015	government	government, research	not declared	open	active

<b>Data source organization</b>	<b>Geographic spread</b>	<b>Taxonomic spread</b>	<b>Temporal spread</b>	<b>Funding origin</b>	<b>Data origin</b>	<b>Data provenance</b>	<b>Access</b>	<b>Activity</b>
ESDAC	global	n/a	n/a	government	government, research	not declared	open	active
FAO	global	agricultural taxa	1961-2020	government	government, other	not declared	open	active
GBIF	global	all taxa	n/a	government, NGO	government, NGO, research	publishers declared	open, licensed	active
GENESYS	global	plant taxa	n/a	government, NGO	government, NGO, research	publishers declared	open, licensed	active
ISO	global	n/a	n/a	government, NGO	government, NGO	not declared	open	active
Landsat/MODIS	global	n/a	1999-2020 (2014-15)	government	government, research	projects declared	open, with registration	active
Map of Life	global	plant/animal taxa	2011-2020	own, research	government, NGO, research	not declared	open, licensed	active
OECD	developed nations (37)	n/a	1961-2020	government	government, other	not declared	open	active
Protected Planet	global	n/a	1981-2020	government, NGO	government, other	not declared	open	active
Soil Grids	global	n/a	n/a	government, NGO	government, NGO, research	not declared	not specified	active
World Economic Plants database	global	plant taxa	n/a	government	government, NGO, research	publishers declared	mostly open, custom terms of use	active
WorldClim	global	n/a	n/a	government, NGO	government, NGO, research	not declared	open, licensed	active



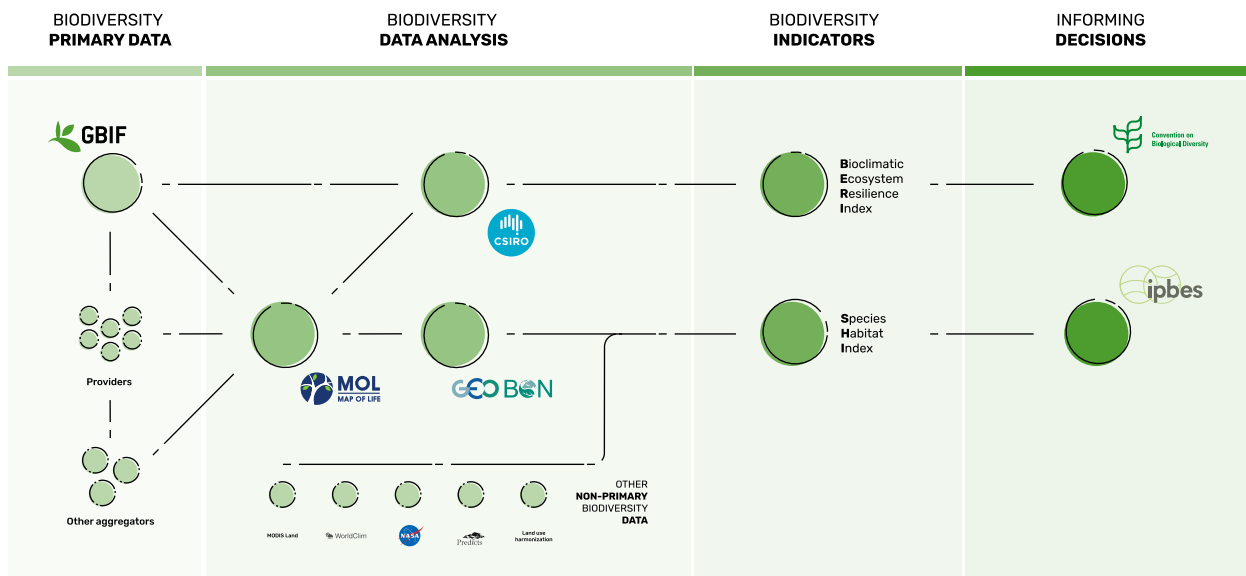
Data source organization	Geographic spread	Taxonomic spread	Temporal spread	Funding origin	Data origin	Data provenance	Access	Activity
WorldGrids	global	n/a	n/a	government, NGO	government, NGO, research	not declared	not specified	inactive
Yale University	global	n/a	1970-2020	government, NGO	government, NGO, research, other	some publishers and projects declared	open	active

3. The pathway for data moving from biodiversity data platforms into the analysis pipeline during the creation of a biodiversity indicator is not always linear. The use or sharing of datasets and data products between indicators magnifies issues of transparency, especially when primary biodiversity data is processed for the benefit of Indicator A and then Indicator A's data products are used as the inputs for analysis for Indicator B. These relationships between indicators are not uncommon. For example, the relationship between the Species Habitat Index (SHI; produced by [Map of Life](#)) and the Bioclimatic Ecosystem Resilience Index (BERI; produced by [CSIRO](#)) demonstrate how data and data product use can become intertwined ([Figure 1A](#)).

The SHI uses biodiversity data from GBIF and other biodiversity data platforms and individual data providers. That data is subjected to various analyses from which Map of Life produces various data products (new datasets) which are then used in part to create the SHI. Similarly, CSIRO takes biodiversity data from GBIF, combines it with data products developed by Map of Life and then uses them to create the BERI. The fact that both use data from GBIF directly demonstrates a likelihood of data overlap, while CSIRO's use of data products from Map of Life that already include GBIF data demonstrates a circular use of data. Adding complexity to this process, both the SHI and BERI utilize data products from the creation of EBVs produced by GEO-BON, which also uses Map of Life data products. The effect of these circular and overlapping data uses essentially creates a 'black box' whose inner workings lack transparency and cannot be discerned easily, if at all. The analysis of the positive or negative effects on the accuracy and effectiveness of a given indicator produced with these types of relationships was not within the scope of this research.

The Agrobiodiversity Index is unique in that it follows a more complex path than other indicators ([Figure 1B](#)), as it builds not only on data and data products but also on other indexes (e.g. the Environmental Performance Index). Transparency becomes more important as the complexity of a given indicator, such as the Agrobiodiversity Index, is increased.

**A**



**B**

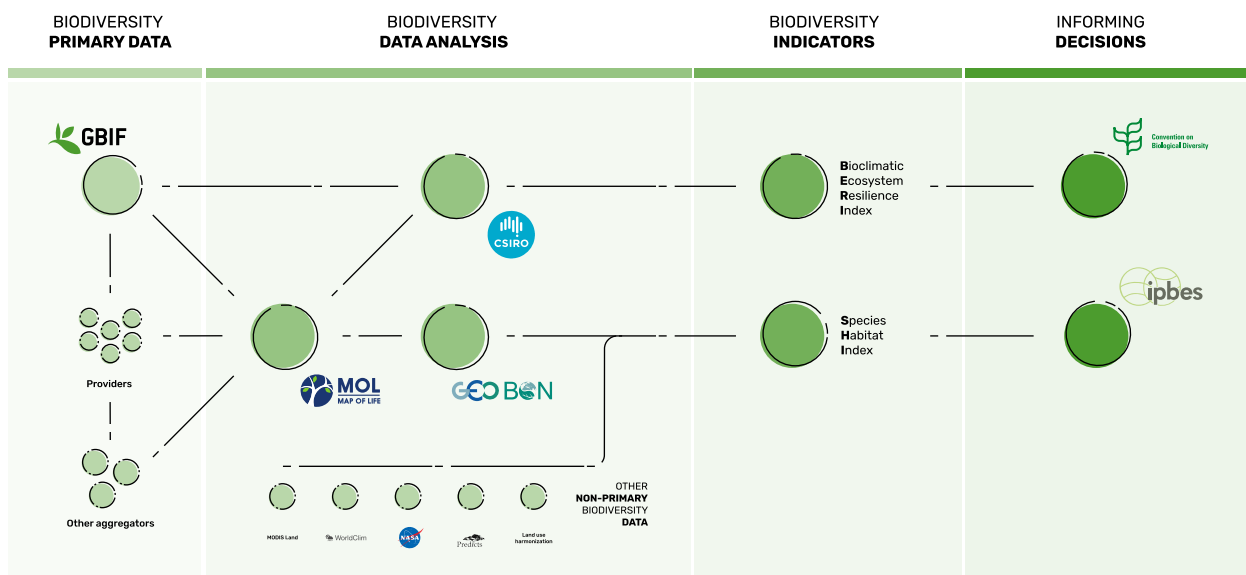


Figure 1. Data workflow / life cycle from data generation, through aggregation or compilation by different sources, to building of biodiversity-related indicators, and dependencies across the distinct organizations involved. **A.** Example for two of the indicators assessed: Species Habitat Index and Bioclimatic Ecosystem Resilience Index. **B.** Example for the Agrobiodiversity Index.

4. Finally, species occurrence data is one of many types of data used to generate EBVs and biodiversity indicators (Table 3). The occurrence data used in these indicators can often be traced back to GBIF, either as direct downloads or as source material for secondary data products produced for EBVs or indicators. The occurrences themselves are derived from multiple sources; they can come from a biodiversity data platform directly (e.g. GBIF); they may be extracted from surveys, inventories, and checklists; and from other maps, peer-reviewed publications, and even from personal research documentation, as demonstrated by the published sources used for the Species Habitat Index, produced by Map of Life. It is worth noting that when more than one platform is used, the result is often the use of shared or duplicate data, such as when data from

GBIF and VertNet (<http://vertnet.org/>) are used (all records in VertNet are also in GBIF). Datasets used in conjunction with species occurrence data encompass a broad range of topics and sources. The use of various forms of geographic data are common, including LANDSAT, MODIS and CGIAR CSI, climatic data (e.g. WorldClim), genetic resources (e.g. GENESYS), and other environmental and agricultural datasets may be used (e.g. SoilGrids; FAO; CIAT; see Table 1).

Table 3. Types of data used for building the biodiversity-related indicators assessed in this study.

Data type

Indicator	occurrence	species info	abiotic info	genetics	other
Growth of Species Occurrence Records Accessible through GBIF	☐	☐	☐	☐	
Species Status Information Index	☐	☐	☐	☐	☐
Species Protection Index	☐	☐		☐	
Species Habitat Index	☐	☐	☐		☐
Biodiversity Habitat Index	☐		☐		
Bioclimatic Ecosystem Resilience Index	☐	☐	☐		
Protected Area Representativeness Index	☐	☐			
Crop Wild Relative Index	☐	☐			
Agrobiodiversity Index	☐	☐			
Comprehensiveness of conservation of socioeconomically as well as culturally valuable species	☐				

## 4. Discussion

The results of this analysis offer insight into the role of primary biodiversity data and other source materials used in the process of generating biodiversity indicators. A review of historic inputs of primary biodiversity data into CBD processes has revealed broad usage of species occurrence data in generating biodiversity indicators. These data are used in conjunction with additional data sources that describe the environmental, genetic, species and geographic characteristics necessary to the focus of the indicator being constructed.

It is clear that biodiversity data platforms, such as GBIF, contribute to the development and use of EBVs and biodiversity indicators by making primary biodiversity data available openly. There remain, however, some challenges within the data and the ways in which the data-sharing community communicates with stakeholders from the local level to national and international policymakers. These challenges present several opportunities to improve the quality of available data and may serve to improve efforts to achieve the goals set in the post-2020 framework.

## 4.1. Relevant and timely high-quality data

Data shared by data platforms and other organizations are critical resources in a wide range of research and scientific endeavors (Ball-Damerow et al. 2019, Graham et al. 2004, Heberling et al. 2021). GBIF, in particular, has played a leadership role in efforts to provide primary data in support of research since its creation in 2001. These efforts must continue in support of research and education globally. To become more relevant to the goals presented by the CBD in the post-2020 framework, however, there are several considerations that may be examined in regard to the types and quality of the data provided through global networks.

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. 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. Although data quality has long been recognized as a key step in the data mobilization process, there is still a large volume of data that still needs to be checked for quality and completeness. As a consequence, 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.

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 and expand their infrastructure to include new data types and formats.

Data quality is an important consideration throughout the process of biodiversity data mobilization. GBIF and other organizations have devoted thousands of hours to train people working with data and to encourage a high standard for data published to public data portals (e.g. BID and BIFA Programmes, VertNet Training Curriculum, iDigBio Workflows). GBIF, for instance, provides a clear set of **minimum requirements for published data**. Once data are published to the GBIF portal, automated processing routines flag data quality issues and produce a set of interpreted fields that use standardized terminology or recommended vocabularies. Such routines help to improve accuracy, increase the discoverability of data and ensure applicability of the data for use across a wide spectrum of research and policy endeavors. While these efforts should continue, they may not sufficiently improve the quality of data to make them useful in the post-2020 framework or EBVs and biodiversity indicators that support it.

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 describe the collection methodology and processes that are as rich as possible. Unfortunately, this type of correction will only be useful for certain types of analyses, and many species occurrence records that represent presence will still not

be likely to be useful for EBVs that must account for absence data. For these EBVs, well-documented monitoring or inventory event data is needed. Certainly, not all data will be useful in this arena, nor are the biodiversity data platforms responsible to make all data high-quality, but some existing legacy data could be enhanced to become more fit for use. Further, data that are currently in the process of mobilization, and all of the data that have the potential to be mobilized, would benefit from an expanded set of best practices for publication.

As more monitoring data becomes available through the network, expanded best practices guidelines should include, but may not be limited to, how to share quality metadata containing details of the sampling methods employed 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. An example of a new standard that is under review for implementation is the Humboldt Core (Guralnick et al. 2017), which would allow more effective sharing of inventory data used for monitoring (Guralnick, pers. comm.).

Currently, many of the datasets published to biodiversity data platforms do not include the metadata adequate to the purpose of addressing bias in the creation of essential biodiversity variables. Several factors contribute to this limitation. In some cases, these data may not have been recorded at the time of collection, while in others it may not yet be digitized from field books and other physical media. For large collections that contain decades worth of data, the process of documenting the metadata for thousands of specimens and collection events may be far too onerous and burdensome. Further, not all data in these collections will have the associated metadata necessary for these improvements, but many institutions will have subsets of their collections that do (e.g. the Grinnell Survey and Re-Survey projects at the MVZ UC Berkeley). To make these data more useful to efforts within the post-2020 framework, it may be worthwhile to encourage institutions to create “sub-collections” for these specific monitoring and inventory projects that they could publish separately from their larger corpus of data. Further, GBIF and other biodiversity data platforms may wish to place greater emphasis on the publication of past and current monitoring and inventory datasets with the expressed purpose of supporting EBV and biodiversity indicator creation. One way to accomplish this might be to strengthen ties with the research and monitoring communities that produce those data and encourage their participation in GBIF’s data publication programmes.

Even if these communities were to satisfy the need for more complete data in different formats, questions would remain about how homogeneous and repeatable the treatment of the same data can be in different contexts. As described above (see Results), the same data, from multiple sources, is being used by distinct organizations or collaborations to build EBVs and indicators, all of which contribute to the same overall goals within the post-2020 framework. Currently, stakeholders developing a given EBV or indicator treat the data independently and apply their own filters and quality checks, which may be more or less similar to those used by other stakeholders. If biodiversity data platforms could prepare species occurrence data in advance for EBV and indicator creation, as EBV-usable datasets, for example, better consistency and transparency might be achieved. In the context of this research, the term “EBV-usable” is meant as described in Kissling et al. (2018a), and refers to datasets compiled from relevant primary data with appropriate licenses that have undergone basic consistency and completeness checks (as opposed to EBV-ready datasets, Kissling et al. 2018a).

This preparation of data presents an opportunity for GBIF to partner with GEO BON and other research and policy organizations to better understand their specific data needs and reinforce their

position as a primary resource for research and analysis-ready biodiversity data. A partnership of this sort might begin with the creation of a community consultation or conversation to identify the key data and data preparations that are needed for EBV generation, including what types of data or metadata are important and what data gaps need to be filled. The results of these consultations could help set future data mobilization goals and targets for GBIF and other data biodiversity data platforms. This conversation might also include identification of community training and capacity-building needs and strategies for addressing them most effectively. For instance, pre-formatted EBV-usable downloads could be provided for EBVs known to need specific subsets of data or datasets in certain formats. Datasets could be prefiltered and preprocessed so that they are readily compatible with other often used ecological or environmental data from data sources such as WorldClim, Landsat, soil grids, Genesys (genetic resources) and geographic regions. These datasets might also be validated against the GBIF/Catalogue of Life taxonomic backbone, which might help to simplify data preparation for EBV creators and remove the need for them to perform additional taxonomic validations.

## 4.2. Support the network

Accomplishing any of the data-related enhancements suggested above will require a concerted effort across the broad biodiversity research and conservation community. The fact that few of the CBD's 2020 goals have shown discernable progress may be attributable, in part, to a disconnect between high-level policy and local activities. In fact, only a few nations have launched organized, country-level monitoring programs that embody shared goals and establish clear lines of communication between policymakers and local practitioners (e.g. national programs in New Zealand (Lee et al. 2005) and Switzerland [Biodiversity Monitoring Switzerland 2021](#); also see [Moussy et al. 2021](#)). This disconnect works in both directions. Calls for data and the organization of large-scale monitoring efforts from entities such as the CBD often do not reach the local communities and organizations that might supply the data, so gaps remain ([Mihoub et al. 2017](#)). Meanwhile, local parties, unaware of these calls from national and international agencies, may not spontaneously contribute data from ongoing efforts, or if aware, they are unable to respond due to the limitations of funding mandates or geographic boundaries ([Johnson et al. 2017](#)).

A plethora of organizations with local foci have made major contributions to the amount of biodiversity data available through data-sharing networks. Much of this data was collected by local volunteers working for self-organized groups attempting to answer questions of local interest ([Schmeller et al. 2009](#), [Pocock et al. 2018](#), [Kühn et al. 2008](#)). While these groups can work in collaboration with one another, they are more often operating on their own outside of a larger national or international framework. Even with a national framework, however, there is no guarantee that a top-down monitoring effort will provide the data or policy results that are needed ([Kühl et al. 2020](#)) without an active network that can link local and national interests and facilitate communication between them.

One of GBIF's most valuable assets is its global network of governments, institutions, and organizations engaged through the efforts of individuals within them. The GBIF network is the social system through which all data flow and the foundation upon which the success of its technical infrastructure resides. This network provides three opportunities that may make biodiversity data from GBIF more relevant to the post-2020 framework, the CBD and new biodiversity goals for 2030 and 2050.



One of these opportunities is the ability 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, which maintain data from the past to the present (Guralnick et al. 2007). As more of these legacy collections have been mobilized, attention has shifted toward monitoring and observation projects, including citizen science (e.g. NOAA's Beach Watch; SANBI's Custodians of Rare and Endangered Wildflowers programme; Chandler et al. 2017) and to biodiversity-focused NGOs (e.g. CERMES; NatureFiji-MareqetiViti). As pressure mounts to address questions about the status and trends of biodiversity, 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).

GBIF is in a unique position to leverage the node-based structure of its network to encourage and train these local agencies and individuals to share monitoring and survey data while applying the best practices for data capture, quality and mobilization. In regions or countries in which a node has not been established, localized institutional networks, nodes from within the same region or nodes from countries with a history of support and involvement for the local effort could contribute. Efforts of this sort have already begun via the BID and BIFA programmes, funded by the European Union and the Government of Japan, respectively. Additional funding and other resources could support and expand these efforts with the explicit purpose of mobilizing local biodiversity data and knowledge. Enabling indigenous peoples and local communities to become active contributors to biodiversity monitoring efforts through the CBD would certainly contribute to meeting the goals set for 2030 and 2050.

Like indigenous peoples and local communities, the private sector is an important source of biodiversity data. Currently, the majority of the data in the GBIF index come from non-profit and government agencies, yet private entities hold a great wealth of biodiversity knowledge in the form of environmental assessments, impact assessments, and other project-based analyses. GBIF has begun to engage with the private sector directly through several initiatives, such as Data4Nature in partnership with the Agence Française de Développement, and the publication of a guidance document to help private companies become publishers through the GBIF network (Figueira et al. 2020). In addition, some national governments have begun to mandate private sector data publication (e.g. Colombia) 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). These developments may present natural opportunities for the GBIF network to support the publication of these data, particularly in countries where a GBIF node exists. In countries or regions without a Node, the GBIF network might provide regional support and private sector help desk, may provide assistance. The CBD would benefit greatly, as would the development of EBVs and indicators that rely on biodiversity data, from a partnership with GBIF and other organizations working to engage the private sector. One key task of this partnership should be to continue to delineate and promote strategies to build upon existing collaborations with the private sector to bring their information holdings into the public sphere.

A second opportunity of equal importance is to mobilize the GBIF network to turn a historically one-way communication pipeline into a more complete cycle. Currently, those organizations and individuals that mobilize data into biodiversity data-sharing portals are hard-pressed to determine

when and how their data are being used by researchers, educators and policymakers. 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 (see [GBIF citation guidelines](#) and the [#CitetheDOI campaign on Twitter and other social media](#)). GBIF is in the uncommon position to be able to continue to build trust across the 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 really 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, if the network does not work toward greater transparency.

The third opportunity for the GBIF network is to mobilize the community to work toward greater transparency and traceability across the entire information supply chain. The creation of EBVs and biodiversity indicators is a complex process. As reported in the Results section, it is not uncommon for the processes and analyses used to generate these 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. Over the last several years, calls have been made to address this lack of transparency ([Navarro et al. 2017](#), [Hardisty et al. 2019](#), [Fernandez et al. in review<sup>^</sup>] [REF?](#)). Each of these calls recognizes that the processes employed and the products produced demand treatments similar to the peer-review process used in academic journals, providing clear documentation, access to primary data and tools for analysis in the short- and long-term. A GBIF partnership with GEO BON, a leader in the facilitation of EBV generation, could be conceived with two key goals:

1. to promote the use of existing guidelines for citation and acknowledgement
2. to improve existing documentation or develop new best practices to accommodate new types and sources of data

Guidelines and best practices that follow FAIR principles and promote the full traceability of EBVs could go a long way toward the realization of the first two opportunities for GBIF network described above. Any partnerships and collaborations should reinforce the [GBIF data users agreement](#) by which all data users are required to “...publicly acknowledge, following the scientific convention of citing sources in conjunction with the use of the data, the Data Publishers whose biodiversity data they have used, where appropriate through use of a Digital Object Identifier (DOI) applying to the dataset(s) and/or data downloads.”

Many individuals in the GBIF network are engaged in the processes of EBV and indicator generation, as well as in those of data acquisition and mobilization. They are likely to be sympathetic to the needs of both groups. With this level of understanding, these individuals should become a key link in promoting mutually beneficial working arrangements between data publishers and EBV and indicator creators. Seizing this opportunity would help both to provide high-quality EBV-useable data and to foster open communication and transparent documentation. If GBIF can take advantage of its relationships with these individuals and groups, the entire community of people in the information

supply chain would profit. Benefits would include receiving recognition for data mobilization efforts and data products created; and an increase in the data traceability, which can improve the reproducibility and transparency of the science. Greater community engagement of this kind would build trust and encourage even greater levels of communication between the policy and research communities.

Data and data products should be assigned persistent identifiers to encourage increased recognition. Currently, GBIF assigns a DOI for every download performed, along with associated information that describes the full query used, including the date and time, number of records, the datasets that contributed to the download, Creative Commons designations and other terms of use and the EML metadata. These data about the query are archived by GBIF indefinitely, but actual downloads are maintained only for a six month period, although data users can request that specific downloads be archived indefinitely. For data sources other than large biodiversity data platforms, identifiers are often missing. In the past, the responsibility to archive or maintain the primary datasets used to create EBVs has fallen upon the organizations in charge of building them. This makes it easy for recognition of GBIF and other data providers to be passed over. Further, it adds to the issues of transparency that make it difficult to replicate the analyses completed for a given indicator because the datasets are no longer available. To remedy this, better communication should be fostered between GBIF, the CBD, BIP, IPBES, GEO BON and other collaborators that promote, build and use indicators, so that archiving of datasets used is secured. GBIF's experience archiving datasets searches and their recent efforts to provide access to monthly snapshots of the GBIF corpus and **derived datasets**, available via cloud services, may increase GBIF's visibility in the CBD and further establish them as an important partner. Ideally, there could be some shared responsibility for these archives that would ensure availability of the data from different access points. Ultimately, a searchable archive of DOIs and associated datasets linked to EBVs and indicators may improve transparency, aid in the reproducibility of the scientific process and improve opportunities for comparisons of data or baselines over time.

## 5. Conclusions

This research has identified four main challenges to the use of primary biodiversity data in the development of EBVs and indicators (described in the Results section), each of which will be critical to the success of the post-2020 global biodiversity framework:

1. Transparency is a critical issue at every stage in the information supply chain. Currently specialized knowledge is required to understand how data flow from data publishers through the processes used to create EBVs and biodiversity indicators.
2. Data sources used in the creation of EBVs and biodiversity indicators are diverse, usually including biodiversity data platforms, with little data coming from the private sector to date. Data sources are generally available for free, although much data is available behind paywalls (e.g. journal access).
3. The pathway from primary biodiversity data to indicator is complicated, the information supply chain is not linear.
4. Data duplication and circular usage is common, yet data traceability from the providers of primary biodiversity data to their use in EBVs and indicators is challenging.

Several actions could help to overcome these challenges:

- An opportunity exists for GBIF to partner with GEO BON and other intergovernmental agencies and NGOs to facilitate further community-wide discussion and continue efforts to identify the key data and data preparations needed to support EBVs, biodiversity indicators, and other related research and policy-focused products. While one-off meetings have occurred in the space, a more persistent working group approach would be more effective. GEO BON has a Data Task Force, but that group has a broad remit and the suggested working group or task force is narrower in application. The new working group might include stakeholders from the scientific and policy communities, as well as a broad spectrum of participants from local-level organizations. This initiative should focus on a series of community consultations hosted by GBIF and GEO BON, with the support of the alliance for biodiversity knowledge, and facilitated by the new working group with a focus on identifying the types of data and metadata needed, the data gaps to fill, and the stakeholders most able to address these needs. Documenting the results of these consultations and a draft set of mobilization goals for community review could refocus existing data collection and mobilization initiatives or guide new ones.
- A second opportunity for a GBIF-GEO BON partnership, along with the **TDWG community**, is the preparation of a set of guidelines in two specific areas:
  - Best practices that support the data collection and mobilization activities identified in the community consultations as described above. These practices would clarify the most effective methods to collect, document, validate, and mobilize data needed for EBV and indicator generation. This process should include individuals and organizations actively engaged in the process of data collection and curation, especially those individuals well-positioned to communicate and promote the use of these best practices in the field.
  - Best practices for building greater transparency and recognition into the processes used to generate EBVs and biodiversity indicators, with a focus on establishing and implementing consistent methods. Community participants should include those individuals and groups responsible for the development of harmonized data products, EBVs and indicators. Additional participants should include members of the data curation community and clearly articulate the importance of transparency and attribution for members of the FAIR data community.
- Data from organized and well-documented monitoring and inventory events are critical for many EBVs and other analyses of trends over space and time. Biodiversity data platforms, such as GBIF, should join ongoing conversations focused on community monitoring activities because of the strong alignment of those models with existing GBIF publishing efforts. A recent paper by Kühl et al. (2020) calls for a model that focuses on opportunistic and semi-structured observations, mapping and strategic long-term data collection at single sites for increased spatiotemporal coverage, all managed by heterogeneous partners that can work together to align their priorities and goals. Another approach could identify new and existing sources of these types of data and to help the community to build capacity to provide those data with high-quality metadata, perhaps using the best practices established by the partnerships described above.
- In conjunction with efforts to mobilize more monitoring data, biodiversity data platforms might provide a series of curated datasets that contain both monitoring and ad hoc species occurrence data that meet a particular standard of quality and fitness for use. This standard could be determined by individual platforms or by the best practices for data collection and mobilization detailed above. These datasets might focus on national and regional scales and should be

archived by one or more organizations publicly (e.g. GEO BON, GBIF). Each dataset should have a DOI assigned and be accompanied by a complete list of credits and attributions.

- As engagement increases between members of the community to establish best practices, new data products and expanded community discussion, it is critical to include indigenous peoples and local communities to facilitate mutually beneficial connections with researchers and policymakers, as highlighted by the UNESCO Recommendation on Open Science (UNESCO 2021). This can take the form of invitations to participate in the consultations described above, but engagement is likely to be limited at best. Instead, an organized effort to engage local and indigenous communities could begin within larger networks, such as GBIF's global network in partnership with organizations with greater experience working with indigenous communities (e.g. CBD, United Nations) framed by the CARE principles. GBIF can play an important role in the adoption of these principles within data workflows and how to highlight mechanisms that support the implementation through its global network. Members of the network who possess knowledge of specific local communities could be enlisted to serve as a conduit between those local individuals and groups and the greater data-sharing community. An individual with local knowledge and ties to the local community in a given country might be identified within the GBIF network. That person could then be supported, financially or through a formal designation, to act on behalf of the broad community to engage with those local groups to begin a conversation about the knowledge possessed within the community and what they would be willing to share. In support of that sharing, the deputy may also learn about the needs of the community and work to establish partnerships with the research and data mobilization communities to build capacity building and provide training necessary to meet local needs.
- The private sector is a source of an exceptional amount of biodiversity data from ad hoc observations and long-term monitoring that is largely unavailable for research and policymaking currently. Some biodiversity data platforms are actively engaging this community to bring their data into open data portals (e.g. GBIF through partnerships such as Data4Nature; UNEP-WCMC Proteus Partnership). As a part of these efforts, GBIF and others should continue to encourage their integration into the larger community conversation, the use of best practices in data collection, and participation in growing initiatives such as the Equator Principles.
- A final action item is the development of a credit and attribution model created by biodiversity data platforms, mobilizers, data users (including organizations such as GEO BON and the CBD that facilitate EBV and indicator creation) and academic journals and other publishers. This collaborative effort should build upon the efforts of groups that have made advances in attribution through the continued use of DOIs and other credit-focused strategies. The growing community of practice focused on data citations includes initiatives such as #CiteTheDOI promoted by GBIF, COPDESS, Make Data Count and a body of literature in support of giving credit where credit is due (Data Citation Synthesis Group 2014, McNutt et al. 2016, Vannan et al. 2020). Journal publishers may hold the greatest leverage to incentivize transparency and recognition, but they will need to be guided by other groups within the community to implement new practices.

## Appendix A: Example

An AsciiDoc appendix is just a document section with a 'specialsection' title.

Place non-essential information that supports your analysis, validates your conclusions or pursues a

related point in an appendix. Examples could include:

- figures/tables/charts/graphs of results
- statistics
- questionnaires
- transcripts of interviews
- survey results
- maps
- software installation instructions

## References

- alliance for biodiversity knowledge (2021) <https://www.allianceforbio.org>
- Ariño AH, Chavan V & Faith DP (2013) Assessment of user needs of primary biodiversity data: Analysis, concerns, and challenges. *Biodiversity Informatics* 8(2): 59-93. <https://doi.org/10.17161/bi.v8i2.4094>
- Ball-Damerow JE, Brenskelle L, Barve N et al. (2019) Research applications of primary biodiversity databases in the digital age. *PLoS ONE* 14(9): e0215794. <https://doi.org/10.1371/journal.pone.0215794>
- Biodiversity Information for Development (BID) Programme. Global Biodiversity Information Facility. Available at:<https://www.gbif.org/programme/82243/>
- Biodiversity Information Fund for Asia (BIFA) Programme. Global Biodiversity Information Facility. Available at:<https://www.gbif.org/programme/82629/>
- Biodiversity Indicators Partnership (2011) Guidance for national biodiversity indicator development and use. Cambridge, United Kingdom: UNEP World Conservation Monitoring Centre. Available at: <https://www.cbd.int/doc/meetings/ind/ahteg-sp-ind-01/other/ahteg-sp-ind-01-bipnational-en.pdf>
- Biodiversity Indicators Partnership - UNEP-WCMC (2021) Growth in Species Occurrence Records Accessible Through GBIF. Available at: <https://www.bipindicators.net/indicators/growth-in-species-occurrence-records-accessible-through-gbif>
- Biodiversity Monitoring Switzerland. (2021). Available at: <https://www.biodiversitymonitoring.ch>
- 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>
- Bush A, Sollmann R, Wilting A et al. (2017) Connecting Earth observation to high-throughput biodiversity data. *Nature Ecology & Evolution* 1(0176). <https://doi.org/10.1038/s41559-017-0176>
- 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>
- Chandler M, See L, Copas K et al. (2017) Contribution of citizen science towards international biodiversity monitoring. *Biological Conservation* 213(B): 280-294. <https://doi.org/10.1016/j.biocon.2016.09.004>



- Collen B & Nicholson E (2014). Taking the measure of change. *Science* 346:166-167. <https://doi.org/10.1126/science.1255772>
- Constable H, Guralnick R, Wieczorek J et al. (2010) VertNet: A New Model for Biodiversity Data Sharing. *PLoS Biology* 8(2): e1000309. <https://doi.org/10.1371/journal.pbio.1000309>
- Data Citation Synthesis Group (2014) Joint Declaration of Data Citation Principles. Martone M, ed. San Diego, Calif., United States: FORCE11. <https://doi.org/10.25490/a97f-egyk>
- Díaz S, Zafrá-Calvo N, Purvis A et al. (2020) Set ambitious goals for biodiversity and sustainability. *Science* 370(6515): 411-413. <https://doi.org/10.1126/science.abe1530>
- 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>
- Essl F, Latombe G, Lenzner B, et al. (2020) The Convention on Biological Diversity (CBD)'s post-2020 target on invasive alien species – what should it include and how should it be monitored? in: Wilson JR, Bacher S, Daehler CC et al. *NeoBiota* 62: 99-121. <https://doi.org/10.3897/neobiota.62.53972>
- Feeley K (2015) Are We Filling the Data Void? An Assessment of the Amount and Extent of Plant Collection Records and Census Data Available for Tropical South America. *PLoS ONE* 10: 1-17. <https://doi.org/10.1371/journal.pone.0125629>
- Figueira R, Beja P, Villaverde C et al. (2020) Guidance for private companies to become data publishers through GBIF: Template document to support the internal authorization process to become a GBIF publisher. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/doc-b8hq-me03>
- GBIF Secretariat (2021) GBIF Science Review 2020. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/bezp-jj23>
- 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>
- Graham CH, Ferrier S, Huettman F et al. (2004) New developments in museum-based informatics and applications in biodiversity analysis. *Trends in Ecology & Evolution* 19: 497-503. <https://doi.org/10.1016/j.tree.2004.07.006>
- Group on Earth Observations Biodiversity Observation Network (GEO BON) (2021). <https://geobon.org/>
- 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>
- Hardisty AR, Michener WK, Agosti D et al. (2019) The Bari Manifesto: An interoperability framework for essential biodiversity variables. *Ecological Informatics* 49: 22-31. <https://doi.org/10.1016/j.ecoinf.2018.11.003>
- Heberling JM, Miller JT, Noesgaard D, Weingart SB & Schigel D (2021) Data integration enables

global biodiversity synthesis. PNAS 118 (6). <https://doi.org/10.1073/pnas.2018093118>

- 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>
- Hoban S, Bruford M, D'Urban Jackson J et al. (2020) Genetic diversity targets and indicators in the CBD post-2020 global biodiversity framework must be improved. *Biological Conservation* 248: 1086542. <https://doi.org/10.1016/j.biocon.2020.108654>
- Hobern D, Baptiste B, Copas K et al. (2019) Connecting data and expertise: a new alliance for biodiversity knowledge. *Biodiversity Data Journal* 7: e33679. <https://doi.org/10.3897/BDJ.7.e33679>
- Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services (IPBES) (2019) Global assessment report on biodiversity and ecosystem services of the Intergovernmental Science-Policy Platform on Biodiversity and Ecosystem Services. Chapter 3: Assessing progress towards meeting major international objectives related to nature and nature's contributions to people. Edited by Brondizio ES, Settele J, Díaz S & Ngo HT. Bonn, Germany: IPBES secretariat. Available at: <https://ipbes.net/global-assessment>
- Johnson CN, Balmford A, Brook BW et al. (2017) Biodiversity losses and conservation responses in the Anthropocene. *Science* 356: 270-275. <https://doi.org/10.1126/science.aam9317>
- Kissling WD, Ahumada JA, Bowser A et al. (2018a) Building essential biodiversity variables (EBVs) of species distribution and abundance at a global scale. *Biological Reviews* 93: 600-625. <https://doi.org/10.1111/brv.12359>
- Kissling WD, Walls R, Bowser A et al. (2018b) Towards global data products of Essential Biodiversity Variables on species traits. *Nature Ecology & Evolution* 2: 1531-1540. <https://doi.org/10.1038/s41559-018-0667-3>
- Kühl HS, Bowler D, Bösch L et al. (2020) Effective Biodiversity Monitoring Needs a Culture of Integration. *One Earth* 3(4): 462-474. <https://doi.org/10.1016/j.oneear.2020.09.010>
- Kühn E, Feldmann R, Harpke A et al. (2008) Getting the public involved in butterfly conservation: lessons learned from a new monitoring scheme in Germany. *Israel Journal of Ecology and Evolution* 54: 89-103. <https://doi.org/10.1560/ijee.54.1.89>
- Lee W, McGlone M & Wright E (2005) Biodiversity Inventory and Monitoring: A Review of National and International Systems and a Proposed framework for Future Biodiversity Monitoring by the Department of Conservation. Landcare Research contract report LC0405/122.
- Lindenmayer DB, Likens GE & Franklin JF (2017) Earth Observation Networks (EONs): Finding the Right Balance. *Trends in Ecology & Evolution* 33(1): 1-3. <https://doi.org/10.1016/j.tree.2017.10.008>
- McNutt M, Lehnert K, Brooks H et al. (2016) Liberating field science samples and data. *Science* 351(6277): 1024-1026. <https://doi.org/10.1126/science.aad7048>
- Mihoub JB, Henle K, Titeux N et al. (2017) Setting temporal baselines for biodiversity: the limits of available monitoring data for capturing the full impact of anthropogenic pressures. *Scientific Reports* 7: 41591. <https://doi.org/10.1038/srep41591>
- Moul K, Martin E, Ivory S et al. (2018) Mapping of current and potential use of biodiversity indicators across intergovernmental processes. Cambridge, United Kingdom: UN Environment

Programme – World Conservation Monitoring Centre. <https://www.unep-wcmc.org/resources-and-data/global-biodiversity-indicator-mapping>

- Moussy C, Burfield I, Stephenson PJ et al. (2021) A quantitative review of species population monitoring. Conservation Biology. Accepted Author Manuscript. <https://conbio.onlinelibrary.wiley.com/doi/10.1111/cobi.13721>
- 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>
- Nicholson E, Collen B, Barausse A et al. (2012) Making Robust Policy Decisions Using Global Biodiversity Indicators. PLoS ONE 7(7): e41128. <https://doi.org/10.1371/journal.pone.0041128>
- 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>
- Pereira HM, Belnap J, Brummitt N et al. (2010) Global biodiversity monitoring. Frontiers in Ecology and the Environment 8: 459-460. <https://doi.org/10.1890/10.WB.23>
- Peterson AT, Asase A, Canhos D et al. (2018) Data Leakage and Loss in Biodiversity Informatics. Biodiversity Data Journal 6: e26826. <https://doi.org/10.3897/BDJ.6.e26826>
- Pocock MJ, Chandler M, Bonney R et al. (2018) A vision for global biodiversity monitoring with citizen science. Advances in Ecological Research 59: 169-223. <https://doi.org/10.1016/bs.aecr.2018.06.003>
- Proença V, Martin LM, Pereira HM et al. (2017) Global biodiversity monitoring: From data sources to Essential Biodiversity Variables. Biological Conservation 213(B): 256-263. <https://doi.org/10.1016/j.biocon.2016.07.014>
- Schmeller DS, Henry PY, Julliard R et al. (2009) Advantages of volunteer-based biodiversity monitoring in Europe. Conservation Biology 23: 307-316. <https://doi.org/10.1111/j.1523-1739.2008.01125.x>
- Schmeller DS, Mihoub J-B, Bowser A et al. (2017) An operational definition of essential biodiversity variables. Biodiversity Conservation 26: 2967-2972. <https://doi.org/10.1007/s10531-017-1386-9>
- Scholes RJ, Mace GM, Turner W et al. (2008) Toward a Global Biodiversity Observing System. Science 321: 1044-1045. <https://doi.org/10.1126/science.1162055>
- Secretariat of the Convention on Biological Diversity (2021) 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 (2020a) Zero Draft of the Post-2020 Global Biodiversity framework. <https://www.cbd.int/doc/c/efb0/1f84/a892b98d2982a829962b6371/wg2020-02-03-en.pdf>
- Secretariat of the Convention on Biological Diversity (2020b) Update of the Zero Draft of the Post-2020 Global Biodiversity framework. <https://www.cbd.int/doc/c/3064/749a/0f65ac7f9def86707f4eaefa/post2020-prep-02-01-en.pdf>
- Secretariat of the Convention on Biological Diversity (2020c) Global Biodiversity Outlook 5. Montreal. <https://www.cbd.int/gbo5>
-

Secretariat of the Convention on Biological Diversity (2020d) Monitoring framework for the post-2020 global biodiversity framework. <https://www.cbd.int/sbstta/sbstta-24/post2020-monitoring-en.pdf>

- Secretariat of the Convention on Biological Diversity (2020e) Review of Progress in the Implementation of the Convention and the Strategic Plan for Biodiversity 2011-2020, document CBD/SBI/3/2 prepared by SCBD for the 3rd meeting of the Subsidiary Body on Implementation (SBI3). Available at: <https://www.cbd.int/doc/c/73bc/335c/480a6a50d95d04478f4b3041/sbi-03-02-en.pdf>
- 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 (2006) Global Biodiversity Outlook 2. Montreal. <https://www.cbd.int/gbo2/>
- Sousa-Baena MS, Garcia LC & Peterson AT (2014) Knowledge behind conservation status decisions: data basis for “Data Deficient” Brazilian plant species. *Biological Conservation* 173: 80-89. <https://doi.org/10.1016/j.biocon.2013.06.034>
- Stevenson SL, Watermeyer K, Caggiano G et al. (2021) Matching biodiversity indicators to policy needs. *Conservation Biology* 35(2): 522-532. <https://doi.org/10.1111/cobi.13575>
- 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>
- Turner W (2014) Sensing biodiversity. *Science* 346: 301-302. <https://doi.org/10.1126/science.1256014>
- UNEP-WCMC & BIP (2020) Indicators for the post-2020 global biodiversity framework. <https://www.cbd.int/sbstta/sbstta-24/post2020-indicators-en.pdf>
- UNESCO (2021) Draft text of the UNESCO Recommendation on Open Science. Available at: <https://en.unesco.org/science-sustainable-future/open-science>
- Vannan S, Downs RR, Meier W et al. (2020) Data sets are foundational to research. Why don't we cite them? *Eos* 101. <https://doi.org/10.1029/2020EO151665>
- Williams BA, Watson JE, Butchart SH et al. (2020) A robust goal is needed for species in the Post-2020 Global Biodiversity framework. *Conservation Letters*. 2020;e12778. <https://doi.org/10.1111/conl.12778>
- Wilkinson MD, Dumontier M, Aalbersberg IJ et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3: 160018. <https://doi.org/10.1038/sdata.2016.18>