Global Biodiversity Data Fitness Assessment

Criteria for decision-grade data

UN environment programme

WCMC

Citation

UNEP-WCMC 2024. Global Biodiversity Data Fitness Assessment: Criteria for decision-grade data. Cambridge, UK: UNEP-WCMC. 6 pp.

Note

This is an updated version of the report by Lauren V. Weatherdon, Bex Gottlieb, Ben Tregenna, Naomi Kingston, Neil Burgess, Corli Pretorius. 2021. *Global Biodiversity Data Fitness Assessment: Progress toward establishing criteria for 'gold standard' decision-grade data.* Cambridge, UK: UNEP-WCMC. 6 pp.

Contact

For further details contact
digitaltransformation@unep-wcmc.org

Cover image

Bird in Costa Rica. (Photo by Zdeněk Macháček, Unsplash)

Acknowledgements

We would also like to thank the One Earth Foundation for supporting the 'Building the Digital Network for Nature' project, and the Swedish Postcode Foundation for their support of the *'Making the next decade count for nature'* (20A712) project, which co-financed this work.

Introduction

A digital ecosystem of nature-related data is crucial for generating the insight required to address the interconnected biodiversity, health and climate crises and to shape global and national policies[1]. To ensure that these 'global public good' datasets are fit-for-purpose, we have developed a methodology for assessing the suitability and 'readiness' of these data for use in informing the evidence base necessary to make informed decisions that create positive outcomes for people and nature.

The aim of this work is to outline progress towards conducting a fitness assessment of global biodiversity datasets, using a set of nine selection criteria to evaluate their fitness for use in supporting progress toward global biodiversity and sustainably frameworks and public and private sector decision-making. These criteria include *policy or decision relevance*, *terms of use*, *accessibility*, *frequency of update*, *temporality*, *geographic coverage*, *transparency*, *scalability*, and *authoritativeness*. The outcomes are intended to assist with prioritizing upgrading efforts and investment in enhancing the fitness-for-purpose of these datasets.

Using these criteria, datasets are classified into three tiers based on their level of alignment, as defined below. Broadly speaking, these tiers can be described as:

- Tier 1: These datasets currently fulfil the criteria to be included as 'global public good' datasets, and added to the World Environment Situation Room, UN Biodiversity Lab and other UN-level platforms.
- Tier 2: These data do not fulfil all criteria. These datasets may be used for assessing global goals, as well as prioritized for investment and upgrading to meet these criteria.
- Tier 3: Datasets that fulfil some of the essential criteria (i.e. authoritative; policy and decision relevant; transparent) and may be prioritized for further development to facilitate their use.
- Not yet qualified: Datasets that do not meet these criteria.

The tiered system is intended to assist in the development and implementation of global datasets that are fit-for-purpose in supporting progress toward the 2030 Sustainable Development Agenda, the Kunming-Montreal Global Biodiversity Framework and related global targets.

[1] See: UNEP. (2019). The Case for a Digital Ecosystem for the Environment. Available at: https://un-spbf.org/wp-content/uploads/2019/03/Digital-Ecosystem-final.pdf.

Assessment criteria and sub-criteria

For each criterion, sub-criteria for assessing alignment are delineated in Table 1. Each of these alignment sub-criteria are allocated a score: 'high/good' = 3; 'medium/moderate' = 2; 'low/poor' = 1; or 'unknown' = data deficient (DD).

Table 1. *Criteria and sub-criteria used to assess each dataset and to get a 'fitness' score assigned.*

	Criteria	Alignment sub-criteria		
		High/Good (score = 3)	Medium/Moderate (score = 2)	Low/Poor (score = 1)
	**1. Policy- or decision-relevance:** Is the dataset *formally listed, recognised or proposed/suitable for use in an MEA for informing and reporting on indicators* (e.g. 2011-2020 Strategic Plan for Biodiversity, proposed in the Kunming-Montreal Global Biodiversity Framework, Sustainable Development Goals, International Finance Corporation Performance Standards or other equivalent global public- or private-sector goal)?	*Formally listed or recognised* for use in informing and reporting on public- or private-sector targets and indicators.	*Proposed/suitable* for use in informing and reporting on public- and private-sector targets and indicators.	*Not formally listed, recognised or proposed/suitable* for use in informing and reporting on public- or private-sector targets and indicators.
	**2. Terms of use:** Is the dataset *open access with minimal restrictions,* while recognising the preferences and rights of data providers to manage access to sensitive information, in accordance with best practices?	Open access dataset available online with minimal restrictions, while recognising the preferences and rights of data providers in accordance with best practices. Examples may include Creative Commons licenses (e.g. CC BY, CC BY-SA) or equivalent, with managed access via APIs where available.	Open access dataset available online with minimal restrictions but requires considerable permissions and paperwork for use. **OR** Dataset available online, but with restricted use (e.g. no use in derivative products; non-commercial use; etc.).	Terms of use unclear or not documented online, preventing informed use of dataset. **OR** Not in accordance with best practices (e.g. misuse of other datasets; does not respect data providers' wishes; no attribution). **OR** Heavily restricted dataset (e.g. not available online; preventing use in derivative products; requiring extensive paperwork).
	**3. Availability / accessibility:** Is the dataset *easily accessible online in different open access, machine readable formats*, including open and free access to associated journal publications and supplementary materials?[2]	Easily accessible (downloadable) online in open access, machine readable formats (e.g. SHP, CSV, KML) and ways to access (e.g. APIs, direct download).	Easily accessible (downloadable) online, but no alternative, open access formats are provided.	Must contact the data provider for access and/or permission.
	**4. Frequency of update:** Has the dataset *been updated regularly (at a rate necessary to be 'current')?*	Regularly updated (at a rate necessary to be 'current'). **OR** Dataset is recognised as the main, up-to-date reference for this feature.	Dataset is out-of-date, but still the only or best dataset for this feature.	Dataset is no longer relevant or is viewed as inaccurate. **OR** Dataset has been superseded by another dataset.
	5. Temporality: Can the dataset *be used to track change over time (where*	≥5 temporal and spatially comprehensive data	3-4 temporal and spatially comprehensive data	≤2 temporal and spatially comprehensive data points available

[2] Note that open, detailed and machine-readable metadata are also important, and are reflected under the '*Transparency*' criterion. A further consideration relates to ontologies—in other words, whether data are linked to existing classification standards (e.g. Darwin Core or industry standards, such as the Global Industry Classification Standard [GICS]) to support interoperability across datasets. Finally, internet bandwidth and other constraints may also affect the accessibility of datasets and should be considered.

	Criteria	Alignment sub-criteria High/Good (score = 3)	Medium/Moderate (score = 2)	Low/Poor (score = 1)
	*applicable),* with a known dataset baseline (three temporal data points minimum, ideally over at least 10 years, e.g., 2010, 2015, 2020)?[3]	points available over the last ten-year period **OR** Not applicable (natural features that do not change on decadal time scales, such as bathymetry or mountains).	points available over the last ten-year period.	over the last ten-year period.
	**6. Geographic coverage:** Is the dataset *globally consistent with comprehensive coverage*, with flexibility to account for national variation while still feasible to be aggregated at a global level?	5+ continents (≥ 20 countries total), with methodological consistency **OR** entire geographic range for feature of interest.	3-4 continents (≥10 countries total) **OR** 5+ continents (<20 countries total) **OR** partial geographic range for feature of interest.	1-2 continents (no matter how many countries included) **OR** 3-4 continents (<10 countries total) **OR** limited geographic range for feature of interest.
	**7. Transparency:** Is the dataset *accompanied by detailed metadata in accordance with global standards (e.g.* ISO 19115, Dublin Core, Darwin Core, and EU INSPIRE), along with a methodology openly and freely published online?	Dataset methodology and detailed metadata are published openly and freely online, in accordance with global standards.	Detailed metadata are published openly and freely online, but the methodology has not been published or is behind a pay wall.	Dataset methodology and detailed metadata are not available online.
	**8. Scalability:** Is the dataset's spatial scale stated and appropriate for upscaling or downscaling to inform national, regional and global decision-making?	Clearly defined spatial scale(s) appropriate for use at national, regional and global scales (or relevant decision-making scale).	Dataset appropriate for use at global scale, but not yet possible to downscale. **OR** Dataset appropriate for use at a national/regional scale, with ongoing efforts toward global coverage.	Spatial scale(s) not defined and/or dataset not fit for purpose for use in decision-making at national, regional or global levels.
	**9. Authoritative:** Has the dataset *been through a peer-review process*, whether published in the scientific literature, reviewed by peers, or a mandated process (e.g. CBD), and is recognised as accurate and authoritative?	Peer-reviewed (e.g. scientific literature, reviewed by peers, or mandated dataset)	Undergoing peer review currently and/or in development.	Not peer reviewed, with no indication of intent.

Scoring

Each criterion is scored on a scale of 1-3, where:

Level	Score range
High/Good:	3
Medium/Moderate	2
Low/Poor	1
Unknown	Data Deficient (DD)

[3] Note that this does not refer solely to the number of releases of the dataset—which is captured under '*Frequency of update*'—but to whether the dataset as a whole has sufficient spatial coverage, methodological consistency, and temporal alignment per release to be able to track change over time of the feature of interest.

To qualify for each tier, the following scores and requirements must be met:

Level	Score range	Further requirements	Description
Tier 1	≥23-27	• Must not have any 'Low/Poor' (1) scoring • Must score 'High/Good' (3) each on *Authoritative* and *Policy-or Decision-relevance*, *Transparency, Terms of Use,* and *Temporality.*	Datasets currently meet the essential criteria fully and should be included as 'global public good' datasets, and added to the World Environment Situation Room, UN Biodiversity Lab and other UN-level platforms.
Tier 2	≥15-22	• Must score 'High/Good' (3) each on *Authoritative, Policy- or Decision-relevance* and *Transparency.*	Datasets can be used for assessing global goals, but do not yet meet all criteria fully. These datasets should be prioritized for investment and upgrading to meet these criteria.
Tier 3	≥12-14	• Must score at least 'Medium/Moderate' (2) each on *Authoritative, Policy- or Decision-relevance* and *Transparency.*	Datasets meet some of the essential criteria and should be prioritized for further development to facilitate their use.
Not yet qualified	0-11	• Score 'Low/Poor' (1) on most or all criteria.	Datasets do not yet meet these criteria but could qualify in the future.