IPBES ICT guide

Part 11 - How to Document an Indicator

Prepared by Joy Kumagai and Aidin Niamir - IPBES Technical Support Unit (TSU) of Knowledge and Data
Reviewed by the Task Force on Knowledge and Data (Hanno Seebens, Rainer Krug, Gregoire Dubois, Matea Vukelić, and Xubin Pan)

For any inquires please contact tsu.data@ipbes.net

This technical guideline discusses how an indicator should be documented within IPBES and is intended for any experts that are involved in creating an indicator following the FAIR and CARE principles. The guideline focuses on the technical aspects of documentation.

I. Introduction

Referenced and used within many biodiversity-related policy documents and instruments, indicators can help us to understand the current state of biodiversity at the ecosystem and species levels, threats and pressures, and conservation responses. They are incredibly important in our current science-policy interface enabling decision making from the international sphere to local responses. As defined by the Biodiversity Indicators Partnership, an indicator is "a measure based on verifiable data that conveys information about more than itself." Indicators often take the form of final datasets resulting from a workflow. An indicator workflow is defined here as the steps performed between the input data and final indicator. Successful biodiversity indicators have a variety of attributes such as scientifically valid, updated over time, responsive to change, easily understood, and consistent as described in the document "Key Knowledge for Successful Biodiversity Indicators (Brooks & Bubb 2014)."

Clear and accessible documentation is necessary for indicators to be effective, consistent, and trustworthy. Unfortunately, indicator workflows often do not follow the FAIR (findable, accessible, interoperable, and reusable) and CARE (collective benefit, authority to control, responsibility, and ethics) principles and are not reproducible by other scientists. For example, small differences in how geospatial layers or assumptions are treated can lead to vastly different results. There needs to be transparency in the underlying calculations, to maintain consistency and ensure reuse by others. IPBES is committed to following the open science principles and increasing transparency in data management. Therefore, this guideline discusses how indicators developed within IPBES should be documented starting from the input data, through processing, to outputs to enable reuse by the community.

Throughout this guideline, we will use the indexes from the publication, Kumagai et al. 2022, as an example for documentation. The data descriptor paper published in Scientific Data describes two indexes which report on how much of six important marine habitats are within protected and conserved areas at the country and global scale

II. Documentation

The recommendations throughout this section focus on practical steps to ensure the FAIR and CARE principles can be followed when documenting the indicator and its calculation.

A. Input data for the calculation of the indicator:

It is essential to document where to find all the necessary inputs that go into the indicator workflow so others can calculate and update the indicator. Therefore, we suggest the following:

- All input data are clearly referenced and noted with version and access date;

- All input data need to be at least findable, although solely using open access datasets in compliance with the FAIR data management principles would be preferable but not feasible in all cases;

- Additionally, all input data used are stored so it is possible to re-calculate the indicator exactly in case the original versions of the data are no longer available;

- If any input datasets are products of Indigenous and local knowledge, it is crucial that a process of Free, Prior, and Informed Consent was completed before the data was generated and the information is used and made available only within the limits and following the intent set by the knowledge-holders;

- The licences of the respective data are acknowledged and followed.

> ⓘ **Example:**
> Following table 2 (shown below) within our example publication, each input dataset is listed together with its reference, the date of access and the version used to calculate the indicator. This information should be associated with the publication of the indicator as, for example, a README file.

Table 2 Description of all data sources used in the workflow.

From: Habitat Protection Indexes - new monitoring measures for the conservation of coastal and marine habitats

Name	Website and Reference	Date of Access	Version
World Database on Protected Areas	Available online at: https://www.protectedplanet.net/en/thematic-areas/wdpa	January 2022	January 2022
World Database on Other Effective Area-Based Conservation Measures	Available online at: https://www.protectedplanet.net/en/thematic-areas/oecms	January 2022	January 2022
Union of EEZs and countries	Available online at: https://www.marineregions.org/[40]	December 2020	Version 3
EEZs	Available online at: https://www.marineregions.org/[41]	March 2021	Version 11
Cold-water Corals	Available online at[42]	January 2022	Version 5.1
Warm-water Corals	Available online at[43]	January 2022	Version 4.1
Knolls and Seamounts	Available online at: https://data.unep-wcmc.org/datasets/41	March 2021	Version 1.0

Mangroves	Available online at[44]	December 2020	GMW 2016
Saltmarshes	Available online at[45]	January 2022	Version 6.1
Seagrasses	Available online at[46]	January 2022	Version 7.1
Ocean	Available online at: https://www.naturalearthdata.com/downloads/110m-physical-vectors/110m-ocean/	December 2020	Version 4.1.0

B. Development and presentation of the indicator workflow:

Ensuring that others can understand the methodology and repeat the workflow of the indicator allows people to improve upon and use the indicators in their own context, greatly expanding the reach and impact of an indicator. Therefore we suggest the following:

- The workflow used to calculate the indicator is contained in reproducible scripts, preferably in an open-source software as explained in Session 4.2 within the IPBES data management tutorials, 'Recommendations for workflow establishment and data management';

- Version control software (such as git behind github) is used to track, store, and share the development as well as the different versions of the workflow, which is especially useful when collaborating in a team. GitHub can not only be used to store the workflow, but the repository can be linked to a Zenodo account (explained here) to make the code citable. Other repositories can also be used. Any new releases of the code should be published with Digital Object Identifiers (DOIs) ensuring traceability;

- A visual figure is created that details the steps for transforming input data to calculate the indicator, in order to help others understand the workflow;

- Documenting the exact versions of the packages, libraries, and software used as well as the operating system is good practice and will help someone to reproduce the exact results by creating the same coding environment. Docker containers and certain packages, such as renv for reproducing packages in R, have been developed to assist with this.

- Any assumptions in the workflow are clearly stated and explained. For example, often when overlaying protected areas with species data, it is assumed that species within those protected areas are "protected," while in reality this is often not the case on the ground. Or another example, often in carbon stock calculations an expert may assume a 1-metre depth of soil.

(i) **Example:**
For the habitat protection indexes, these recommendations were followed while developing the workflow. The workflow was created using open source software, R, instead of ArcGIS Pro as this ensured consistency in processing the data and there is no requirement to buy a licence to use the workflow. All scripts used to process the original datasets and to calculate the indexes can be found within the associated GitHub folder, linked respectively within the data availability and code availability sections of the publication. To enable reusability, the workflow is also structured in a user-friendly way so that the entire process can be run using one R script. Finally, figure 5 (shown below) within the publication provides an overview of the workflow to explain what each script does and how it fits into the overall process of creating the indexes.
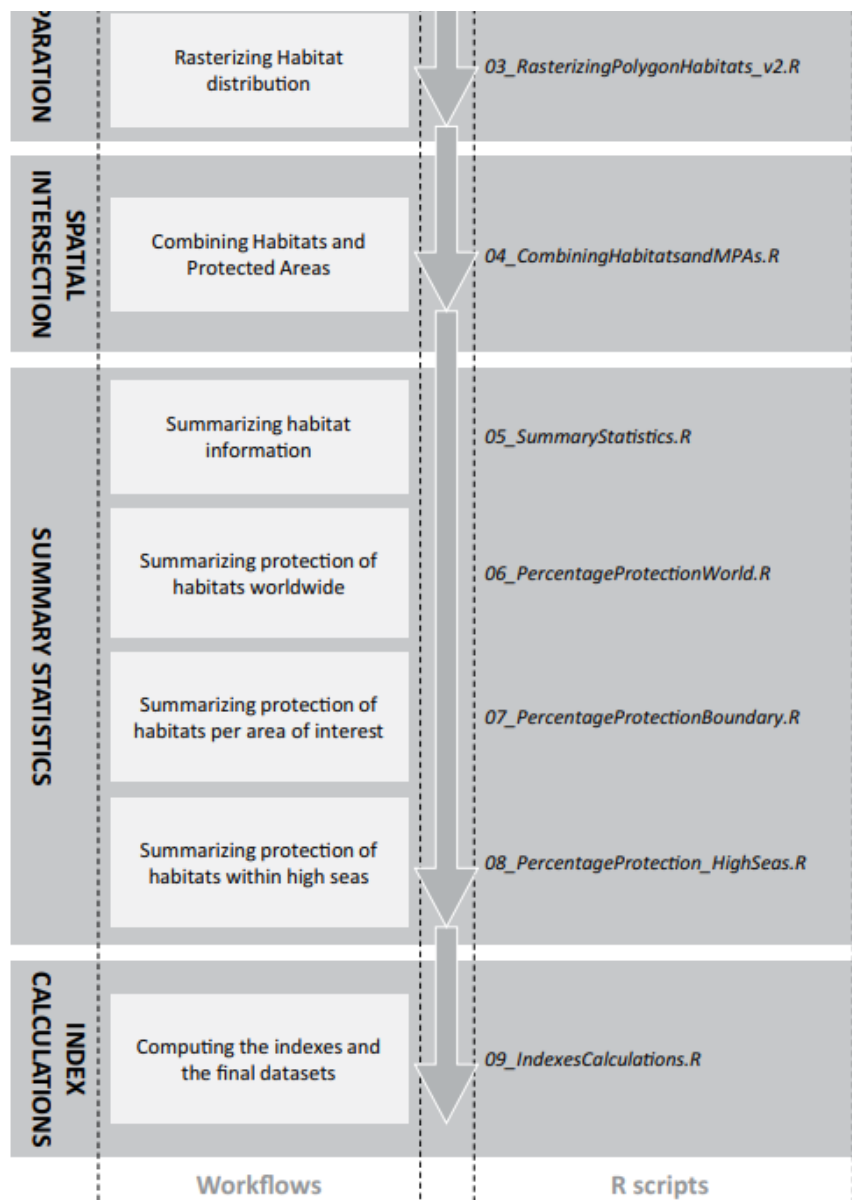
Fig. 5 A flow chart describing the key steps of the indexes calculations. We also connect each step to the R script available at: https://github.com/jkumagai96/Marine_Habitat_protection where a more detailed explanation on how to replicate the workflow is available.

C. Outputs

Now that the inputs and process behind the indicator are clearly documented and explained, the final dataset (i.e. the indicator) should be findable and accessible. Additionally, in the documentation there should be information about the interpretation and limits, e.g. the scale and time frame in which the indicator can be interpreted. Therefore, we recommend the following:

- The final indicator dataset is easily findable and accessible in an online repository and has been assigned a DOI;
- Within this repository, the indicator dataset is stored in an interoperable format with clear metadata (version number, temporal scale, geographic scale, etc.);
- A licence is assigned to the indicator dataset and associated workflow, preferably CC-BY or CC0;
- A contact person is clearly identified in the repository with a monitored email;
-
  The indicator is communicated and interpreted clearly, including its limits and what questions it was

developed to address, in a report/publication/media release to share it with the wider community;

- The workflow used to create the indicator is also available within a repository that contains a readme file (see here) explaining the inputs, processing steps, and how they work together, and licence so future users can re-create the indicator. As mentioned in B, GitHub is a great option for storing the workflows behind the indicator, and can be used with Zenodo to make the code citable by publishing releases of final versions of code with a DOI ensuring traceability (explained in detail here).

> ⓘ **Example:**
> These guidelines are followed within our example.
>
> - The final indicator dataset is provided in an online and open repository on Zenodo (https://doi.org/10.5281/zenodo.4694821). The final datasets are provided as CSV files (an interoperable file format), and the repository has clear metadata including a DOI, versioning, contact information, and authors' list (see image below).
>
> - The linked publication and affiliated media reports serve to communicate the indexes to a broader audience.
>
> - The workflow scripts are also stored in the same repository with a README file (https://github.com/jkumagai96/Marine_Habitat_protection) that explains the inputs, coding environment, and the file structure to be able to repeat the workflow.

In general, the associated publication should provide all required information to reproduce the results, including a link to the workflow repository and final dataset.

zenodo

March 4, 2022   Dataset   Open Access

Habitat Protection Indexes - new monitoring measures for the conservation of threatened marine habitats - Datasets and supporting files

ⓘ Joy A. Kumagai; ⓘ Fabio Favoretto; ⓘ Sara Pruckner; Alex D. Rogers; ⓘ Lauren V. Weatherdon; Octavio Aburto-Oropeza; ⓘ Aidin Niamir

The supporting datasets, scripts, and supplementary information for the manuscript, "Habitat Protection Indexes - new monitoring measures for the conservation of threatened marine habitats," are available within this repository.

We conduct an analysis on the coverage of protected areas that cover six threatened marine and coastal and developed two indexes, the Local Proportion of Habitat Protected Index and the Global Proportion of Habitat Protected Index, describing the protection of these habitats locally and globally. The habitats considered are the following: cold corals, warm water corals, knolls and seamounts, mangroves, saltmarshes, and seagrasses.

The index scores of each jurisdiction are made available for download in the dataset: *habitat_protection_indexes_average.csv*

The habitat specific index scores for each jurisdiction are made available for download in the dataset: *habitat_protection_indexes.csv*.

Column name descriptions are available in the text file: *Column_name_descriptions_20220301*

Column name descriptions are available in the text file: *Column_name_descriptions_20220301*

The scripts used to run the workflow to calculate the indexes, create figures, and calculate statistics for the manuscript are also included. The script *01_Workflow sources* the first 9 scripts in the *scripts* folder to calculate the indexes which relies on the functions script within the functions folder. The rest of the scripts in the folder create the figures and calculate the statistics for the manuscript.

A readme pdf file is included here to ease with reproducing the workflow, but we strongly suggest to please visit our github (https://github.com/jkumagai96/Marine_Habitat_protection) to reproduce the entire calculation where we provide detailed information on how to run the workflow and package management.

III. IPBES Process

Within IPBES, it is important that if any expert team develops an indicator, the associated technical support unit informs the technical support unit on knowledge and data. The technical support unit on knowledge and data can ensure that these steps are followed and all input data are accessible and stored in case the original source of the information is removed, enabling reuse by future assessments. Increasing transparency in the process of creating an indicator ensures consistency, allowing the scientific community to repeat workflows, improve upon them, and test different scenarios. Over time this process will build trust for the indicator and encourage their use to help inform policy and action.

Your feedback on this content is welcome. Let us know what other useful material you would like to see here by emailing tsu.data@ipbes.net.

References

1. Brooks, S. and Bubb, P. (2014) Key Knowledge for Successful Biodiversity Indicators UNEP-WCMC, Cambridge, UK 12pp.
   https://www.bipindicators.net/system/resources/files/000/000/410/original/901.pdf?1482313832

2. Kumagai, J.A., Favoretto, F., Pruckner, S. et al. (2022) Habitat Protection Indexes - new monitoring measures for the conservation of coastal and marine habitats. Sci Data 9, 203.
   https://doi.org/10.1038/s41597-022-01296-4