

How-To

This page provides insight into the EBV netCDF data format, offering support and guidance for your upload process.

1. [Why is netCDF the right format for EBVs?](#)
2. [What are the key features of the EBV netCDF files?](#)
3. [What metadata are included?](#)
 - a. [What is the General Information needed to document your data?](#)
 - b. [What are the EBV attributes and how do they impact the data cube structure?](#)
4. [What are entities?](#)
5. [What are EBVs?](#)
6. [What is a good title?](#)
7. [Where can I find inspiration to start?](#)
8. [Why is iDiv always the publisher in the dataset citation?](#)

1. Why is the EBV netCDF structure the right format for storing Essential Biodiversity Variables?

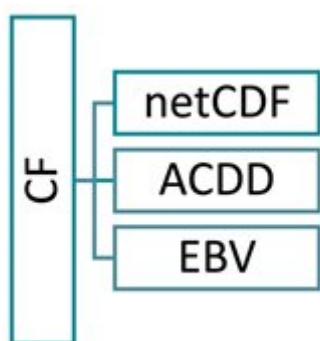


Figure 1: Core elements of the EBV

netCDF

© Langer/ iDiv

The EBV netCDF structure is designed to hold Essential Biodiversity Variables (EBV). The EBV concept is further described [here](#). The files are based on the [Network Common Data Form](#) (netCDF). Additionally, it follows the [Climate and Forecast Conventions](#) (CF, version 1.8) and the [Attribute Convention for Data Discovery](#) (ACDD, version 1.3).

The Network Common Data Form was chosen because of its ability to accommodate the hierarchical, multidimensional data format outlined in the subsequent section. Furthermore, the CF convention provides the foundation for the internal structure of the EBV netCDFs.

This includes e.g., the definition of the dimensions, coordinate variables, georeferencing and more, thereby enabling to use these files with common GIS software such as GDAL and QGIS. Also, the CF convention is increasingly used in the community of Earth sciences data. This convention covers the core metadata terms. These metadata terms are enriched by the ACDD. Additional terms for the precise description of EBV specifications were added. The result is a self-describing data set holding raster data for Essential Biodiversity Variables.

2. What are the key features of the hierarchical data structure in the EBV netCDF files?

The structure allows several data cubes per netCDF file. These cubes have **four dimensions**: longitude, latitude, time and entity, whereby the last dimension can, e.g., encompass different species or groups of species, ecosystem types or other. Each cube holds data of a specific metric.

The usage of hierarchical groups enables the coexistence of multiple data cubes, each sharing the same dimensions. The first hierarchical level (netCDF groups) are scenarios, e.g., the modelling for different Shared Socioeconomic Pathways (SSP) scenarios. The second hierarchical level (netCDF groups) represent metrics, such as the percentage of protected area per pixel or the proportional loss over a certain time span per pixel. In contrast to the scenario-level (which is not mandatory), each EBV netCDF must have at least one metric. All metrics are repeated per scenario, if any are present. The number of scenarios and metrics included in the data sets can and will vary.

Just keep in mind: All EBV netCDF files always have a metric. But they may or may not have a scenario at the higher level. The resulting data cube(s) hold(s) the data per metric. These data cubes are 4D.

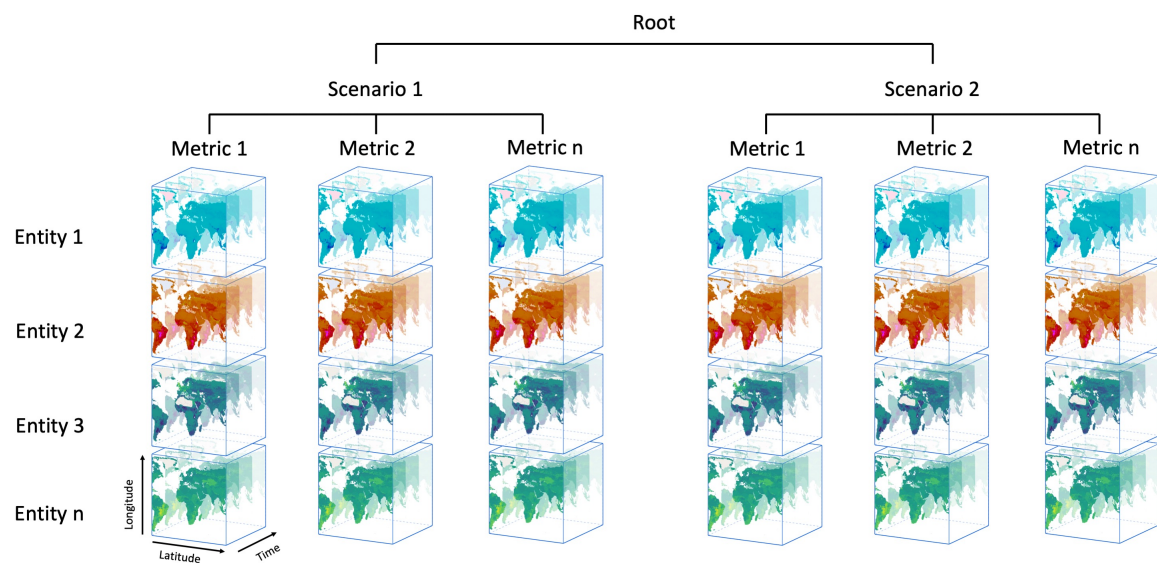


Figure 2: EBV netCDF hierarchical structure
© Langer/ iDiv

3. What metadata terms are included and how do they relate to the upload form?

The metadata in the netCDF file is based on the information that you provide in the upload form here in the portal. The following tables bundle these terms. In the netCDF files itself there is more metadata than you can find here in this tables. There are technical terms which are added for, e.g., the noData-value description, the georeferencing and more.

As the names in the upload form differ from the ones in the netCDF, both are shown in the table. Remember: the metadata terms in the netCDF (mostly) follow a convention. That's why the name in the upload form sometimes differs: to make it clearer for the users. The mandatory attributes are indicated by the little star.

In addition to the terms listed below, the name, email address and institution of the publisher are automatically added to the metadata. The publisher is the person who submits the data set to us for review. If you want another person to be the publisher, you can transfer the rights to another person in the "My uploads" section of the "Upload" tab before submitting the data set for review.

a. What is the General Information that is needed to document your data?

Name: upload form	Name: netCDF	Description
Title*	title	A short phrase or sentence describing the data set.
Date of creation*	date_created	The date on which this version of the data was created in YYYY-MM-DD format.
Summary*	summary	A paragraph describing the data set, analogous to an abstract for a paper.
References	references	Published or web-based references that describe the data or methods used to produce it.
Methods*	source	The method of production of the original data. If it was model-generated, source should name the model and its version. If it is observational, source should characterize it.
Coverage Content Type*	coverage_content_type	Describes the source of the data based on an ISO 19115-1 code (image, thematicClassification, physicalMeasurement, auxiliaryInformation, qualityInformation, referenceInformation, modelResult, or coordinate).
Project Name	project_name	The name of the project(s) principally responsible for originating this data. Multiple projects can be separated by commas.
Project URL	project_url	The URL(s) of the project(s).
Creator Name*	creator_name	The name of the person principally responsible for creating this data.
Creator Email	creator_email	The email address of the person principally responsible for creating this data.

Creator Institution*	creator_institution	The institution of the creator; should uniquely identify the creator's institution.
Co-creators	contributor_name	The name of any individuals, projects, or institutions that contributed to the creation of this data.
License*	license	Provide the URL to a standard or specific license.

b. What are the EBV attributes and how do they impact the data cube structure?

On the second page of the upload form, you are asked for specific attributes related to the EBV class and name, entity, metric, scenario, spatial scope, and temporal resolution of your data set.

A key information is bundled in the metric and the scenario section. It is important to keep in mind, that even though, they are next to each other, inside the netCDF file these are nested hierarchical. Meaning: If you, e.g., add two scenarios and one metric, the latter will be in both scenarios which means you will have two data cubes – one for each scenario-metric combination. The metrics are basically all the data sets you have for one EBV. So, if you have the absolute number of birds (count) and additionally the change over time (percentage), you have to add two metrics.

You will also need to fill in the information regarding the entity, like the entity type and scope. Remember, the entity is one of the dimensions of the data cube(s). For example, the entity type could be 'Species' and the scope could be 'African great apes'. The actual entities are the seven African great apes – you need to give this information as a CSV table on the last page of the upload form. Check out some examples to understand this concept in more depth (see [section four](#) of this page).

EBV class and EBV Name:

Name: upload form	Name: netCDF	Description
EBV Class*	ebv_class	EBV Class of the data set. (The EBV Class is GEO BON's classification system for biodiversity monitoring. It categorizes and organizes essential ecological variables across scales and biological levels. This standardized framework helps identify and prioritize key variables representing biodiversity aspects like species composition, population dynamics, ecosystem functioning, and habitat quality.)
EBV Name*	ebv_name	EBV Name of the data set. (The EBV Name is a unique identifier for a specific variable in the EBV Class, representing a distinct aspect of biodiversity. It helps identify and categorize measured or monitored biodiversity information. The EBV Name reflects the ecological attribute being assessed, such as "Species Richness," "Population Abundance," "Functional Diversity," "Habitat Fragmentation," or "Ecosystem Productivity".)

Entity attributes:

Name: upload form	Name: netCDF	Description
Entity Type*	ebv_entity_type	EBV entity type, e.g., 'Species'.
Entity Scope(*)	ebv_entity_scope	Specifies the entity scope in more detail, e.g., 'African great apes'.
Classification System Name	ebv_entity_classification_name	Name of the classification system used for the entity types (optional).
Classification System Reference	ebv_entity_classification_url	URL of the classification system used for the entity types (optional).

Metric attributes:

Name: upload form	Name: netCDF	Description
Name*	standard_name	A short name of the metric.
Description*	long_name	A longer description of the metric.
Units*	units	The units of the variable's data values.

Scenario attributes

Name: upload form	Name: netCDF	Description
Name*	standard_name	A short name of the scenario.
Description*	long_name	A longer description of the scenario.

Spatial scope

Name: upload form	Name: netCDF	Description
Spatial scope*	ebv_spatial_scope	Spatial scope of the data set, either 'Continental/Regional', 'National', 'Sub-national/Local' or 'Global'.
Spatial description(*)	ebv_spatial_description	Specific information about the spatial scope.

Temporal resolution:

Name: upload form	Name: netCDF	Description
Temporal resolution	time_coverage_resolution	Describes the targeted time period between each value in the data set (ISO 8601:2004 date format).
Temporal extent	time_coverage_start & time_coverage_end	The time span covered by the data set.

Environmental domain:

Name: upload form	Name: netCDF	Description
Environmental domain	ebv_domain	Environmental domain of the data set, one or several of 'Terrestrial', 'Marine' or 'Freshwater'.
Comment	comment	Miscellaneous information about the data, not captured elsewhere.

4. What are entities?

We already introduced the 'entity' as the fourth dimension of the EBVCube above. This is the technical point of view.

From the biological perspective an entity can encompass very different scopes. It can for example be a certain species, so you have entities at the species level. Also, an entity could bundle a group of species, such as, forest birds and non-forest birds. Very common is also the existence of ecosystem type based entities. Simply said: the different entities that you have data for (per metric).

In regards of the upload process, you need to upload a simple *.csv file with the names of the entities. These names get stored in the EBV netCDF and connected to the entity-dimension. These names should be human-readable. That means, please don't provide, e.g. ecosystem type codes or abbreviations, but instead the full name that is directly understandable. The csv itself should simply have one column, each row one entity, no header.

5. What are EBVs?

If you have ever wondered why biodiversity changes and why it matters, you're in the right place. Essential Biodiversity Variables (EBVs), provide a comprehensive framework that helps us understand and protect variety of life on Earth. EBVs can be imagined as a set of standardized and harmonized observations, specifically designed to prioritize and uncover biodiversity change. They capture the essential dimensions and organizational levels of biodiversity, allowing us to study, report, and manage these changes effectively. Just like Essential Climate Variables monitor the climate, EBVs enable us to integrate diverse data from various sources, allowing us to monitor biodiversity across different spatial scales – from your local neighborhood to our entire planet – at the national, regional, or even global level. But EBVs go beyond just data collection. By linking these measurements to meaningful indicators, we gain insights that inform policies aimed at protecting and conserving biodiversity. Join us on this journey; together, we can make a difference and ensure that biodiversity data evolves into valuable information. Through the power of EBVs, we gain meaningful insights that empower us to make informed decisions and take action to safeguard our natural heritage.

6. What is a good title?

When choosing a title for your EBV dataset, follow these guidelines:

1. Be clear and specific in conveying the focus and content of your dataset.
2. Strive for brevity while maintaining informative and meaningful title.
3. Use descriptive keywords highlighting important elements or variables in your EBV dataset.
4. Ensure abbreviations are consistent throughout the title and in the extended EBV dataset description.
5. Tailor your title for the target audience, consider the intended users and adapt the title accordingly.
6. Avoid ambiguity:
 1. Only include specific methods, models, metrics, or indicators if it helps disambiguation.
 2. Only include spatial and temporal information for context if it helps disambiguation.
 3. Only indicate the geographic scope if it helps disambiguation.
7. If your data got published before try to connect the title to this publication.

Note that the title should serve as a teaser for your dataset and not replace the full metadata. A well-crafted EBV dataset title should provide enough information to captivate the user's interest and encourage further exploration of the details!

7. Where can I find inspiration to start my upload process?

We generally encourage you to look at the other data sets in the portal for some inspiration. If you are not sure what to fill in, you can also hover over the (blue) highlighted text on the upload form to see an example. You can also always contact us via ebvportal@idiv.de for all additional questions.

Tips for example datasets: If you have data for **one metric ('only')** and **no scenario** you can check out [the Habitat availability for African great apes](#) by Jessica Junker (by the way, this data set has seven entities – one per great apes species). If you have **at least two metrics but no scenario**, you can take a look at the [Nocturnal bird migration across western Europe](#) by Raphaël Nussbaumer (this data set has only one entity: Nocturnal birds. The more eye-catching part of this data set is its immense temporal dimension: 322 dates). If you modelled your **data for different scenarios**, e.g., the SSP scenarios the [BES-SIM cSAR-iDiv](#) by Inês Martins is a good example to follow (this data set has three entities based on different species-groups: non-forest birds, forest birds and all birds).

To explore the data, you can always click on the 'Show on map' button on the detail page for each data set. That allows you to directly see the data in a map and check out the scenarios, metrics and entities of the data set without downloading.

To discover the EBV netCDFs in full detail, we recommend [Panoply](#). This is a software developed by NASA to generally open netCDFs. This allows you to really see the internal hierarchy and it also has a plot function. This can be a bit overwhelming as you also see all the distributed technical components and attributes. But it is a very nice way to deeply understand the EBV netCDFs without code.

You can also download and open the EBV netCDFs directly in your R code with the [ebvcube package](#). This package bundles all the important metadata for you and hides all 'unnecessary' technical stuff. Also, you can directly start working with the data. It provides some easy high-level functions to, e.g., directly visualize or subset the data as you like.

8. Why is iDiv always the publisher in the dataset citation?

The publisher in the dataset citation is based on the DOI. The DOI is provided by DataCite. From the DataCite-perspective, iDiv is always the publisher. This set-up is represented by the netCDF attribute *naming_authority* which is defined as '*The organization that provides the initial id for the dataset. [...]*' (ACDD 1.3). The DOI publisher is different from the publisher field in the EBVCube metadata. The publisher field in the EBVCube metadata is an ACDD term and contains '*The name of the person [...] responsible for publishing the data file or product to users, with its current metadata and format.*'