Building Essential Biodiversity Variable netCDFs with the ebvcube R Package

Luise Quoß[‡,§], Néstor Fernández[‡,§], Christian Langer[‡,§], Jose Valdez[‡,§], Miguel Alejandro Fernández [‡,§], Henrique M. Pereira[‡,§]

‡ German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany
§ Institute of Biology, Martin Luther University Halle Wittenberg, Halle, Germany

Corresponding author: Luise Quoß (luise.quoss@idiv.de)

Abstract

The concept of Essential Biodiversity Variables (EBVs) was conceived to study, report, and manage biodiversity change. The EBV netCDF structure was developed in order to support publication and interoperability of biodiversity data. This standard is based on the Network Common Data Format (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 standard allows several datacubes per netCDF file (see Fig. 1). These cubes have four dimensions: longitude, latitude, time and entity, whereby the last dimension can, for example, encompass different species or groups of species, ecosystem types or other aspects. The usage of hierarchical groups enables the coexistence of multiple EBV cubes (see Fig. 2). The first level (netCDF group) are scenarios, e.g., the modelling for different Shared Socioeconomic Pathways (SSP) scenarios. The second level (netCDF group) are metrics, e.g., the percentage of protected area per pixel and its proportional loss over a certain time span per pixel. All metrics are repeated per scenario, if any are present. The result is a rather complex raster dataset (see example dataset in Fig. 3).

This is where the ebvcube R package comes into play. This R package enables scientists to create their own netCDFs in the EBV cube standard. Its functionality covers the creation, opening/reading and visualizing the EBV netCDFs. The ebvcube package is part of the overall EBV infrastructure and works together with the EBV Data Portal. Users can work with the downloaded EBV netCDFs or upload their own EBV netCDFs to the portal.

Generally, the package aims to condense the output for the users and assist in the understanding of the file structure to overcome the complexity. The output is reduced to the necessary information, e.g., not displaying coordinate variables or any technical attributes. Moreover, functionality for a quick data exploration is implemented.

Keywords

GEO BON, data standard, EBV, monitoring, interoperability, FAIR, data portal

Presenting author

Luise Quoß

Presented at

TDWG 2022

Conflicts of interest

References

- Mahecha M (2017) Earth System Data Cube. figshare. Figure. https://doi.org/10.6084/m9.figshare.4822930.v2

Figure 1.

Abstract visualization of the EBV cube (Mahecha 2017, License CC BY 4.0).

```
ebv-dataset                                      ebv-dataset
  └─global attributes                              └─global attributes
  └─dimensions [lon, lat, time, entity]            └─dimensions [lon, lat, time, entity]
  └─(coordinate) variables                         └─(coordinate) variables
  └─metric_1                                       └─scenario_1
     └─group attributes                               └─group attributes
     └─ebvcube [lon, lat, time, entity]               └─metric_1
        └─datacube attributes                            └─group attributes
                                                         └─ebvcube [lon, lat, time, entity]
                                                            └─datacube attributes
                                                      └─metric_2
                                                         └─...
                                                      └─...
                                                   └─scenario_2
                                                      └─...
                                                   └─...

  (a)                                              (b)
```
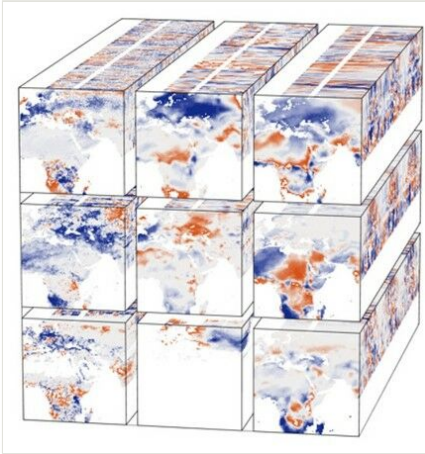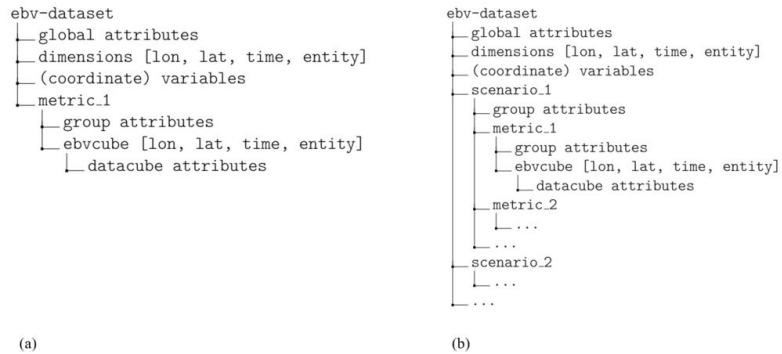
Figure 2.

EBV netCDF hierarchical data structure: (a) shows the structure of a minimal dataset and (b) shows the structure of an exhaustive dataset.

```
mammals dataset
 ├─title:  InS1GHTS_LUH2_2015_2055
 ├─summary:  Data on Area Of Habitat (AOH) for 100
 │  mammals from 2015 to 2055, in 5 year intervals
 ├─ebv_class:  Species populations
 ├─ebv_name:  Species distributions
 ├─ ...
 ├─dimensions [lon, lat, time, entity]
 ├─lat
 │  └─ lat attributes
 ├─lon
 │  └─ lon attributes
 ├─time
 │  └─ time attributes
 ├─entities
 │  └─ entities attributes
 ├─crs
 │  └─ crs attributes
 ├─scenario_1
 │  ├─standard_name:  Sustainability
 │  ├─long_name:  SSP1-RCP2.6
 │  ├─metric_1
 │  │  ├─standard_name:  Habitat availability
 │  │  ├─long_name:  absolute values per 5 years and
 │  │  │  species - in km2
 │  │  ├─ebvcube [720,1 440, 9, 100]
 │  │  │  ├─units:  land-use of mammals calculated in km2
 │  │  │  ├─grid_mapping:  crs
 │  │  │  └─ ...
 ├─scenario_2
 │  ├─standard_name:  Middle of the Road
 │  ├─long_name:  SSP2-RCP4.5
 │  ├─metric_1
 │  │  ├─group attributes
 │  │  └─ ...
 ├─scenario_3
 │  └─ ...
 ├─scenario_4
 │  └─ ...
 ├─scenario_5
 │  └─ ...
```

Figure 3.

EBV netCDF Structure of the Global habitat availability for mammals dataset by Daniele Baisero (License CC BY 4.0).

Blue elements are variables, green elements represent groups, attributes are displayed in black and dimension are red.