Measurement Ontologies: A Field of Dreams for Essential Biodiversity Variables

Margaret O'Brien[1], Mark Schildhauer[1], Steven Chong[2], and Robert Miller[2]

[1]University of California, Santa Barbara
[2]University of California Santa Barbara

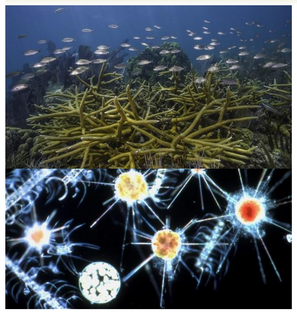
November 24, 2022

Abstract

Essential Biodiversity Variables (EBVs) are state variables that lie between primary measurements and high-level indicators, and are necessary for assessment of the health and prognosis of Earth's biosphere. EBVs represent the complete spectrum of biological diversity from genes to ecosystems, and so are based on observations which themselves are highly diverse, and typically human-collected or analyzed. What is now sorely needed are structured dictionaries of biological measurements that data collectors, curators and nascent biodiversity programs can reference at all stages of planning and data organization. Similarly, analysts working with data defined according to these measurement dictionaries, require assurance that their results are comparable across scales and institutions. Full understanding of primary measurements will ideally require machine-readable, interpretable, and interoperable descriptions of the measurement contents, collection methods, data-typing, dimensions and associated units for physical quantities, and specification of appropriate temporal and spatial scales, plus the relationships among those attributes and facets of the ecosystem. Formal ontologies, i.e. vocabularies built using modern Semantic Web technologies, now provide the ideal tools and protocols for structuring and operationalizing EBV primary measurements. Here we illustrate an approach to apply these to existing data sets (both primary and harmonized intermediates) using community-accepted measurement ontologies under development. Such techniques can streamline the discovery and integration of observations, assist with calibration/validation checks required for automated or remote data collection, and enable rigorous structured definitions for modeled or remotely-sensed EBVs as these are developed.

Posted on 24 Nov 2022 — CC-BY 4.0 — https://doi.org/10.1002/essoar.10500833.1 — This a preprint and has not been peer reviewed. Data may be preliminary.

American Geophysical Union 2018, B41L-2896



Measurement Ontologies: A Field of Dreams for Essential Biodiversity Variables

Margaret O'Brien[1,2], Mark Schildhauer[2], Steven S Chong[2], Robert J Miller[1]

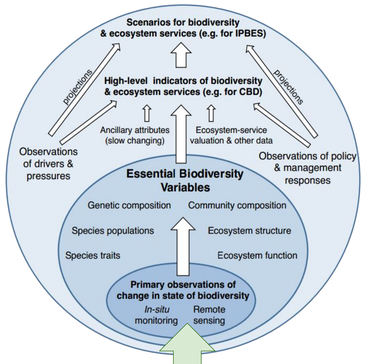
[1]Marine Science Institute & [2]National Center for Ecological Analysis and Synthesis, U.C. Santa Barbara

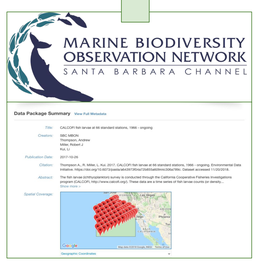
[contact: mobrien@ucsb.edi; ORCIDs - mob:0000-0002-1693-8322; ms:0000-0003-0632-7576; ssc: 0000-0003-1264-1166; rjm:0000-0002-8350-3759]

Essential Biodiversity Variables

Biological data are collected at great expense, with the expectation that they will **improve our capacity for science-based decision-making,** aimed at *protecting* natural ecosystems and *sustaining* the services that they provide.

The **Marine Biodiversity Networks (MBON)** strengthen our understanding of marine biodiversity through scientific observations that form the basis of Essential Biodiversity Variables (EBV)[1].





EBVs characterize critical aspects of ecosystems, and are intended to function as *interfaces* between *raw data*, and *indicators* of biodiversity change[2]

(EXAMPLE EBV)

EBV Class: *Ecosystem Function*

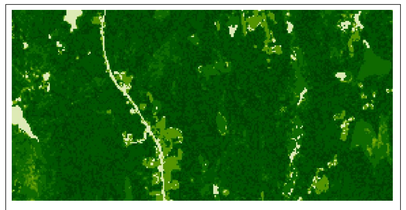
EBV Candidate: *NPP- Net Primary Production*[3]

Measurements of carbon fixation and flux are collected in many different ways, as determined by scope and context. This measurement heterogeneity creates challenges in estimating NPP

in situ chamber measures NPP on a rocky reef

Many Chamber and remote-sensed datasets estimating NPP are publicly available, with rich metadata & values in units of "(Carbon) mass per area per time"

However, most measurements of NPP are not readily harmonized



Satellite image for estimating NPP from space

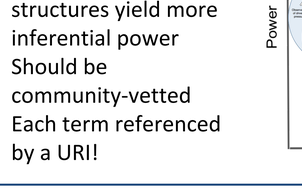
Ontologies Clarify Content of Primary Observations

"Measurement Types"

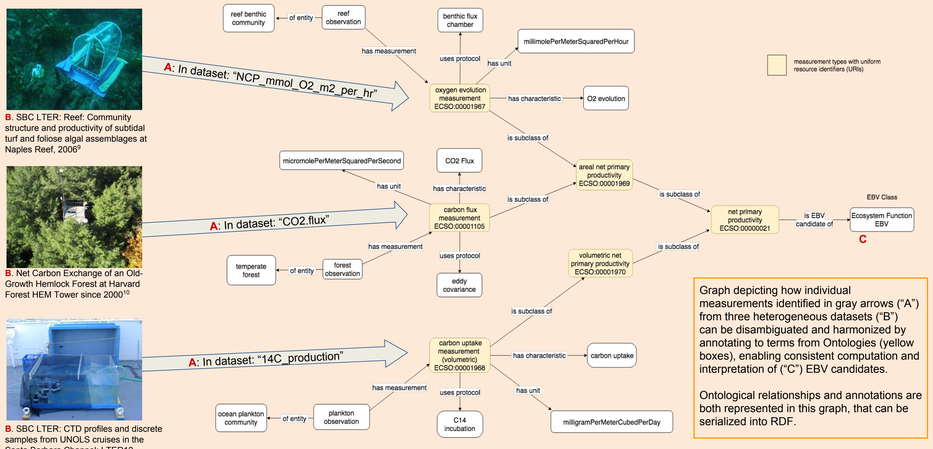
Harmonizing raw data through an observational framework

Ambiguity exists in what the raw data represent (i.e. the actual observation and measurements), as well as in the "definitions" of the EBV's. Ontologies can help clarify their meaning and contents

Construction of "Measurement Types" in ECSO (using OBOÉ pattern):



- Use W3C recommended languages (OWL, RDF) to develop formal ontologies of concepts/terms (e.g. NPP) and their relationships (e.g. to CO2 flux)
- Useful ontologies constructed using OWL (http://obofoundry.org)--
  - **ENVO**: Environment Ontology - terms for biomes, environmental features and materials, e.g., boreal moist forest, ocean pelagic zone; soil, water, air, carbon
  - **PATO**: Phenotype And Trait Ontology - terms for organism qualities, properties, attributes or characteristics, e.g, red, ectopic, high temperature, fused, small, edematous, arrested
  - **PCO**: Populations and Communities Ontology - entities, qualities, and processes related to collections of interacting organisms,
  - **CHEBI**: Chemicals of Biological Interest - molecular entities, 'small' chemical compounds
  - **BCO**: Biological Collections Ontology - processes related to specimens and occurrences
  - **UO**: Unit Ontology - Metric dimensions and units (established to work with PATO)
- Use "Observations and Measurements" Design Pattern to represent how terms (from above) are actually recorded in ***Raw Data.***
  - **OBOE**: Extensible Observation Ontology - links entities-of-interest, characteristics, units, protocols, context. Interoperable with OGC's Observations and Measurements (O&M), and interactions of sensors, observation, sampling and actuators (SOSA)
- **ECSO ontology** expresses "Measurement Types" that describe measurements in the raw data using OBOÉ-- capturing both the ***Entity of Interest***, as well as the ***Characteristic measured*** using terms from ENVO, PATO, CHEBI, BCO, PCO, and other OWL ontologies



B. SBC LTER: Reef: Community structure and productivity of subtidal turf and foliose algal assemblages at Naples Reef, 2006[9]

B. Net Carbon Exchange of an Old-Growth Hemlock Forest at Harvard Forest HEM Tower since 2000[10]

B. SBC LTER: CTD profiles and discrete samples from UNOLS cruises in the Santa Barbara Channel: LTER12, 2005-01-20 to 2005-01-27[11]

Graph depicting how individual measurements identified in gray arrows ("A") from three heterogeneous datasets ("B") can be disambiguated and harmonized by annotating to terms from Ontologies (yellow boxes), enabling consistent computation and interpretation of ("C") EBV candidates.

Ontological relationships and annotations are both represented in this graph, that can be serialized into RDF.

Binding Data to Ontology Concepts

With *semantic annotation*, datasets should reference community-accepted measurements whenever possible. *Applications can access the complete description of a term from an Ontology via its URI*

Metadata (EML):
URI reference to ontology term embedded in EML attribute metadata, indicating a measure of EBV "Population Abundance"

```
<attribute>
  <attributeName>pctcov</attributeName>
  <attributeLabel>percent cover</attributeLabel>
  <attributeDefinition>The percent ground cover on the field</attributeDefinition>
  ...
  <annotation>
     <propertyURI label="contains measurements of type">ECSO:containsMeasurementsOfType</propertyURI>
     <valueURI label="Plant Cover Percentage">http://purl.dataone.org/odo/ECSO_00001197</valueURI>
  </annotation>
</attribute>
```

Darwin Core Archive: measurement URIs can populate the measurementID field, here using a defined measurement for volumetric abundance, describing EBV, "Population Abundance"

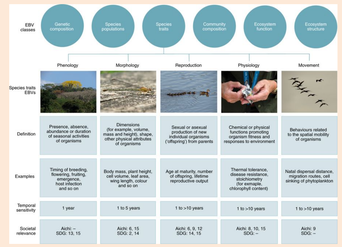
eventID	date	Taxon	taxonID	measurementOrFact	measurementID	measurementValue	measurementUnit
80116_1_5m	16-Jan-2008	Pseudo-nitzschia	ITIS:584561	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	5960	numberPerLiter
80116_1_5m	16-Jan-2008	Leptocylindrus	ITIS:2394	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	3040	numberPerLiter
80116_1_5m	16-Jan-2008	Thalassiosira	ITIS:2484	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	2680	numberPerLiter
80116_1_5m	16-Jan-2008	Chaetoceros	ITIS:2758	cells_liter	http://purl.dataone.org/odo/ECSO_00001201	4740	numberPerLiter

Metadata (ISO 19115):
URIs would appear in code lists, referenced in the column descriptor. Ideally, URI's refer to terms defined in OWL Ontologies

```
<gmd:MD_Band>
  <gmd:sequenceIdentifier>
    <gco:MemberName>
      <gco:aName>
        <gco:CharacterString>biomass</gco:CharacterString>
      </gco:aName>
      <gco:attributeType> {6 lines}
    </gco:MemberName>
  </gmd:sequenceIdentifier>
  <gmd:descriptor>
    <gco:CharacterString>Biomass</gco:CharacterString>
  </gmd:descriptor>
  <gmd:units xlink:href="http://unitsofmeasure.org/ucum.html#kg"/>
</gmd:MD_Band>
```

Complementary work - Species Traits

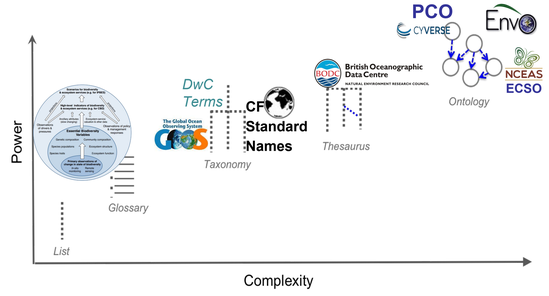
Kissling et al[3] (2018, Fig 2., right) suggest five EBVs within the class 'species traits' and define their measurements, temporal sensitivity and societal relevance. Traits selected for EBVs (e.g., body mass, plant height) allow quantification of species' response to global change at the population level. **EBV-relevant Traits will be defined in ONTOLOGIES to make them clear and comparable**



Disambiguating heterogeneous data through Controlled Vocabularies

Controlled Vocabularies:

- Fall on a spectrum
- More complex structures yield more inferential power
- Should be community-vetted
- Each term referenced by a URI!



OWL Ontologies:

- Describe classes, instances, properties, and relationships
- Provide consistent machine- readable interpretations

References, Links and Datasets

1. Perreira et al. 2013. doi: 10.1126/science.1229931
2. Kissling et al. 2018. doi: 10.1111/brv.12359
3. Kissling et al. 2018. doi: 10.1038/s41559-018-0667-3
4. https://geobon.org/ebvs/what-are-ebvs/
5. ENVO: http://environmentontology.org/
6. PATO: http://www.obofoundry.org/ontology/pato.html
7. PCO: http://www.obofoundry.org/ontology/pco.html
8. OBOE: https://github.com/NCEAS/oboe
9. Dataset doi: 10.6073/pasta/a627d0c9a79fd625894956d6f19c49e0
10. Dataset doi: 10.6073/pasta/e3cf11c0810f12b2e6a8996dbdc45117
11. Dataset doi: 10.6073/pasta/59c6aeff5dfdb39f32dfad1db432bffc
12. Dataset doi: 10.6073/pasta/a643973f04a72b855a60944c306a789c
13. Dataset doi: 10.6073/pasta/d302929b97723a1425364e1a19efbf55
14. Dataset doi: 10.6073/pasta/be54fff488ec4d0d6146956b5fff057e

http://marinebon.org, https://mbon.ioos.us/