

GEO Essential Variables (EVs) as a basis for a common semantic framework to better integrate heterogeneous in-situ data sources into the European Green Deal Data Space



Ivette Serral (ivette@creaf.uab.cat)¹, Joan Masó¹, Anthony Lehman², Raul Palma³
¹ CREAF, Autonomous University of Barcelona, Spain
² Institute for Environmental Sciences (ISE), University of Geneva, Switzerland
³ Poznan Supercomputing and Networking Center (PSNC), Poznań, Poland

AD4GD project has received funding from the European Union's Horizon 2020 Research and Innovation Action Programme under GA 101061001. Co-funded by the European Union, Switzerland and the United Kingdom. September 2022 - August 2025

Green Deal Data Space

All Data 4 Green Deal. An Integrated, FAIR Approach for the Common European Data Space
 Co-create and shape the European Green Deal Data Space as an open hub for FAIR data and standards-based services that support the key priorities of pollution, biodiversity and climate change



The common European Green Deal Data Space (GDDS) will interconnect currently **fragmented** and **dispersed data** from various ecosystems, both for/from the private and public sectors, to support the objectives of the European Green Deal. It will offer an **interoperable**, trusted IT environment for **data processing**, and a set of rules of legislative, administrative and contractual nature that determine the rights of **access to and use** of the data.

The **European Green Deal. Accessible and interoperable data** are at the heart of data-driven innovation. This data, combined with **digital infrastructure** and **artificial intelligence** solutions, facilitate evidence-based decisions and expand the capacity to understand and tackle environmental challenges.

How to operate and discern among all the data available in the GDDS?



FAIR principles semantics

There are many vocabularies for variables and observations, but a formalization is lacking. AD4GD proposes to use **Essential Variables**.

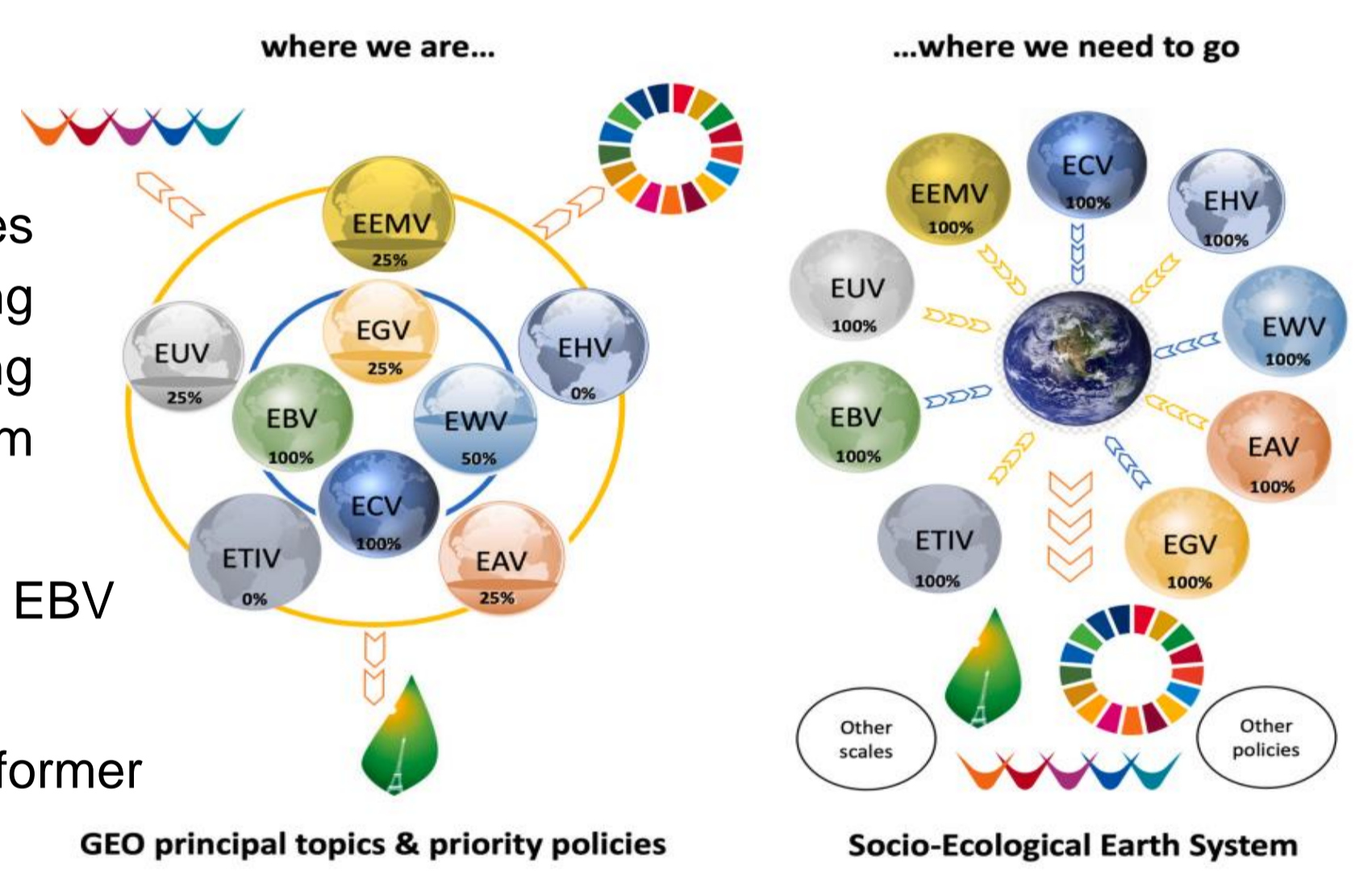


Essential Variables (EVs)

Essential Variables (EVs) have been defined by different communities as sets of variables that are crucial for characterizing and monitoring systems across space and time, providing insight into underlying processes and their changes, and/or feeding indicators that inform environmental policies at multiple scales.

The most consolidated EVs are ECV (Climate), EOV (Ocean), EBV (Biodiversity), EWV (Water) and EAV (Agriculture).

Within GEO they are being extended to other thematic communities (former SBAs) and formalized through the GEO Community Activity on EVs.



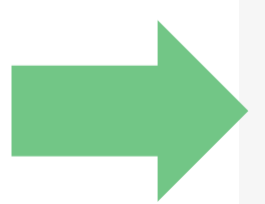
Extracted from: Essential earth observation variables for high-level multi-scale indicators and policies. A. Lehmann et al. Environmental Science and Policy 131 (2022) 105–117

AD4GD proposal for EVs semantic tagging

All EVs are being encoded in **OGC RAINBOW** definition server so they can be referred as a vocabulary with a permanent and unique identifier.



OGC RAINBOW is intended to be a node in an interoperable ecosystem of resources published by different communities.



Vocabulary

Preferred Label: **Essential Biodiversity Variables - codelist scheme**

URI: <http://w3id.org/ad4gd/ev/ebv/>

Definition: Essential Biodiversity Variables (EBVs) are defined as a minimum set of measurements, complementary to one another, that can capture major dimensions of biodiversity change. EBVs are organized in six classes (Genetic composition, Species populations, Species traits, Community composition, Ecosystem functioning, Ecosystem structure) and cover the three realms (Marine/coastal, Terrestrial and Freshwater)

Concept Hierarchy:

- Community composition
- Ecosystem Structure
- Ecosystem function
- Ecosystem structure
- Genetic composition
- Species populations
- Species traits

Description: Essential Biodiversity Variables (EBVs) are defined as a minimum set of measurements, complementary to one another, that can capture major dimensions of biodiversity change. EBVs are organized in six classes (Genetic composition, Species populations, Species traits, Community composition, Ecosystem functioning, Ecosystem structure) and cover the three realms (Marine/coastal, Terrestrial and Freshwater)

Identifier: EBV

is Defined By: <https://geobon.org/ebvs>

Object Type: <http://www.w3.org/2004/02/skos/core#ConceptScheme>

Preferred Label: Essential Biodiversity Variables - codelist scheme

See Also: <https://www.opengis.net/def/entities/bodies/ogcna>

Source: <https://www.opengis.net/def/entities/bodies/ogcna>

Title: Essential Biodiversity Variables - codelist scheme

Collections: Concepts in Essential Biodiversity Variables - codelist scheme

Data coming from observations need to link to this vocabulary.

Sensor Things API+ OGC standard is perfect for in-situ observations as it has mechanisms to connect with semantics through the observedProperty.

BUT WHAT ABOUT OTHER DATA FORMATS WITHOUT METADATA SEMANTIC TAGGING such as CVS?

In **AD4GD** we propose a similar idea of the STA+:

- A **file** (extended JSON schema, CSVW, etc) to describe the semantic tagging.
- A **service** that will store these files without having to modify the original data when this is not possible: the **meaning service**.
- With this, it's possible to **automatically import** this file into a STA+ service.
- Sensor data in interoperable STA+ is then ready to ingest the **GDDS**,

