Proposal to extend the I-ADOPT Framework to support the aggregation of I-ADOPT variables

Gwenaëlle Moncoiffé, Barbara Magagna, Sirko Schindler, Robert Huber, Petra Ten Hoopen, Alison Pamment, Anusuriya Devaraju, Maria Stoica

March 2023

One important use case of the I-ADOPT framework is the necessity to link I-ADOPT variables to multiple broader concepts. This is not only to facilitate their discovery by a range of end-users across a broad variety of applications, domain specialism and/or language registry, but also to support aggregation of variables in data products.

Many examples for this requirement can be found in various research domains. Below are some emerging ones from the environmental domains:

- Essential Ocean Variables (EOV)[1], are groups of variables used for the quantification of key ocean processes, selected for their relevance, feasibility and cost-effectiveness. Grouped by topics, they are described in the Global Ocean Observing System (GOOS)'s specification sheets[2]
- Essential Climate Variables (ECV)[3], are groups of variables used to characterise the Earth's climate also grouped by topics and layer
- Essential Biodiversity Variables (EBV)[4], are groups of variables classified into species and ecosystem-focused themes
- eLTER Standard Observations (SO)[5], classifications of variables into different categories of the Ecosystem Integrity concept

In order to categorize I-ADOPT variables under one or multiple grouping concepts, it is necessary to extend the I-ADOPT framework adding a new class for the grouping concepts and four new relations, as follows:

VariableSet

> *An aggregation class to group a set of variable for a specific purpose*
> *Example: EOV "**Nutrients**", ECV "**Precursors for Aerosols and Ozone**", EBV "**Ecosystem function**", example*

hasApplicableProperty

> *Domain: VariableSet Range: Property   Cardinality 1..n*
> *A VariableSet has one or more applicable Properties*

---

[1] https://doi.org/10.1111/gcb.14108

[2] https://www.goosocean.org/index.php?option=com_oe&task=viewDoclistRecord&doclistID=168

[3] https://public.wmo.int/en/programmes/global-climate-observing-system/essential-climate-variables

[4] https://doi.org/10.1038/s41559-019-0826-1

[5] https://elter-ri.eu/storage/app/uploads/public/62c/ea2/a00/62cea2a002845239798196.pdf

*Example: The EOV "[inorganic carbon](#)" may regroup variables with the following applicable I-ADOPT properties: "**alkalinity**", "**pH**", "**concentration**" (of dissolved inorganic carbon), "**partial pressure**" (of carbon dioxide)*

hasApplicableObjectOfInterest

Domain: VariableSet   Range: ObjectOfInterest   Cardinality 1..n
A VariableSet has one or more applicable ObjectOfInterest.
*Example: The EOV "[nutrients](#)" may regroup variables with the following chemical substances as applicable "object of interest": **nitrate**, **nitrite**, **ammonium**, **phosphate**, **silicic acid (or silicate)**.*

hasApplicableMatrix

Domain: VariableSet   Range: Matrix   Cardinality 0..n
A VariableSet has zero or many applicable Matrices.
*Example: The EBV "Ecosystem function" may regroup variables with measurement of primary productivity in different environmental domains or matrices (e.g. **in sea water**, **in sediments**, **in ice**)*

hasApplicableContextObject

Domain: VariableSet   Range: ContextObject   Cardinality 0..n
A VariableSet has zero or many applicable ContextObjects.
*Example: The SO "[Soil inventory](#)" may regroup variables with measurements in the **geosphere** as context object*

ro:hasMember (http://purl.obolibrary.org/obo/RO_0001901)

Domain: VariableSet   Range: Variable   Cardinality 1..n
A VariableSet has or many Variables as members.
Example: The SO "[Soil inventory](#)" may regroup variables like amount of organic matter, soil bulk density, soil pH value.
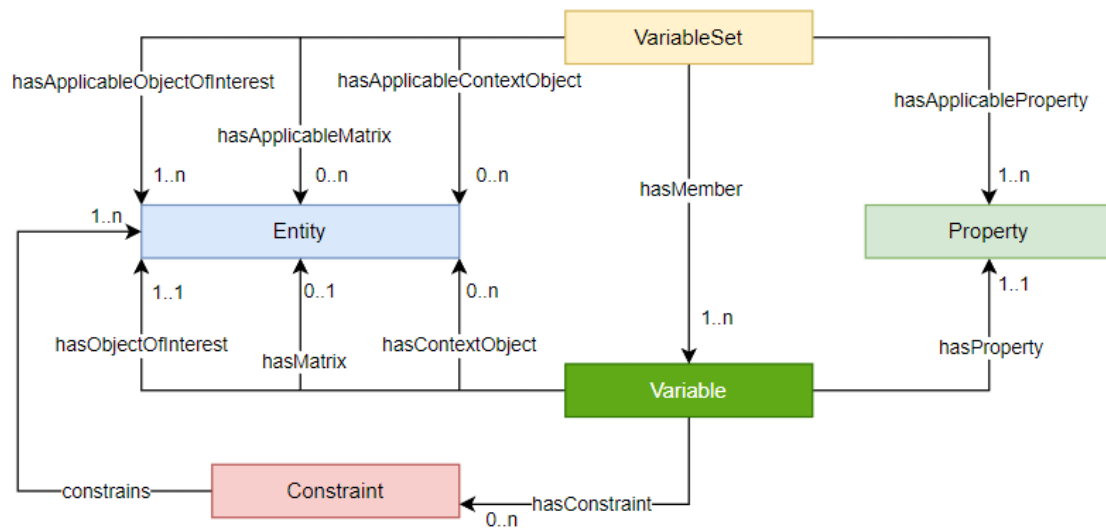
Figure1: Extended I-ADOPT Framework by VariableSet and associated relations

For example of use cases that may apply the proposed extension, see appendix.

Appendix:

Adoption case: British Oceanographic Data Centre (BODC)

BODC is working on an implementation of the I-ADOPT ontology that seeks to connect Essential Ocean Variables (EOV) to variable names described in, for example, the [BODC Parameter Usage Vocabulary (PUV)](#) or the [Climate and Forecast (CF) Standard Names](#) collections.

This is needed in order to construct demonstrators in projects such as ENVRI-FAIR and EOSC-future, to enable a user to select an EOV, for example "Oxygen", from a drop-down list, and for search engines to return only the datasets that contain variables that are relevant to that selected EOV from a range of participating Research Infrastructures (RI).

Two approaches could have been taken:

1- we could have asked every RI to tag their datasets with the relevant EOV keyword "Oxygen" and repeat this for every EOV and every datasets.

2- or we could connect the concept "Oxygen" to the variable naming schemes used by the RIs.

There are many advantages in using the second approach. First it builds capacity for future interoperability across resources by encouraging RIs to use variable naming conventions that are aligned with international community standards. Second, it builds capacity to easily link variables to other types of aggregating terminologies (in a "plug-and-play" manner). Indeed, the linkage between the aggregating term and the variable name could be static ("hard-coded") through a semantic link between the URLs of the 2 concepts, or it could be dynamic through a SPARQL query containing the logic that will associate the aggregating term to the relevant variable names. The maturity and well known grammar or logic used to construct the CF Standard Names and the BODC PUV parameter codes, and the availability of the I-ADOPT framework gave us the opportunity to develop the second approach to connecting concepts. We started referring to this as "smart mapping". This is because it does not require the maintenance of many static mappings and it also gives us the flexibility of creating complex aggregations on demand, adapting the logic of the reasoner to what our users want.

Our first step was to decompose a selection of CF Standard Names and PUV parameter codes according to the I-ADOPT ontology classes and to link the components to instances within a common set of controlled vocabularies (one for the chemical object of interest, one for the property observed, one for the matrix in which the measurement was made) using the I-ADOPT ontology properties iop:hasProperty, iop:hasObjectOfInterest, iop:hasMatrix.

We then started doing the same with the EOV concept (using instances from the [AtlantOS Essential Variable vocabulary A05](#)) until we realised that this was breaking the I-ADOPT rules. Indeed, in many cases, an aggregated concept needed to be linked to multiple applicable "properties" or multiple applicable "Objects of Interest" (i.e. all the ones that qualified).

We realised that this was clearly not an acceptable use of the IOP ontological properties and we invited input from others via the I-ADOPT github repo: https://github.com/i-adopt/questions/issues/1 .

 See examples below.

Example 1:
http://vocab.nerc.ac.uk/collection/A05/current/EV_CO2/?_profile=iop&_mediatype=text/html

Example 2:
http://vocab.nerc.ac.uk/collection/A05/current/EV_NUTS/?_profile=iop&_mediatype=text/html

In the first example, it can be seen that EV "Inorganic Carbon" (defined here: https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=17475) would need to harvest datasets containing pH or total alkalinity measurements as well as partial pressure of CO_2, or dissolved inorganic carbon concentrations in water bodies. So it needs to connect to a range of I-ADOPT-compatible variables but it cannot itself be defined as an I-ADOPT variable because it regroups observations that belong to different quantity kinds with different dimensions. This is incompatible with the cardinality definitions of the existing properties and classes of the I-ADOPT ontology.

In the second example, the term "Nutrients" is a collective term that regroups a number of distinct chemical entities and would therefore need to be mapped to more than one iop:ObjectOfInterest. The concept "nutrients" also means different things for different communities and contexts. So each mapping between the broad concept "nutrients" and observed chemical substances is highly context dependent. For the GOOS EOVs collection, "Nutrients" consist of nitrate, nitrite, phosphate, and silicate (https://www.goosocean.org/index.php?option=com_oe&task=viewDocumentRecord&docID=17474). But the concept "nutrient" taken in isolation could be mapped to different combinations of chemical substances depending on context. Again this required us to avoid hard mappings and adopt a flexible/customisable approach. We believe that by creating a dedicated class for the grouping concept and dedicated relationships that mirror those already defined for the I-ADOPT variables we could solve this important issue.

Adoption case: European long-term ecosystem, critical zone and socio-ecological systems Research Infrastructure (eLTER)

eLTER is working on the representation eLTER Standard Observations (SO) based on the eLTER's whole system approach (WAILS) that can characterise adequately the state and future trends of the Earth systems. In an harmonisation effort, the SOs were developed to provide most critical information from the diverse primary observations in a standard format and in this sense they can be compared to Essential Variables.SOs also consider aspects of cost-effectiveness and operative feasibility as preconditions in the consideration of methods and protocols applied to the variables measured.

eLTER Standard Observation represents a set of minimum variables. Each variable is described with a specific method and protocol. In total around 70 SOs for geosphere, hydrosphere, atmosphere and biosphere with several hundred variables were identified.

Each variable will be represented via the decomposition approach of I-ADOPT in the EnvThes Vocabulary. In order to be reusable for the Dynamic Ecological Information Management System - Site and dataset registry (DEIMS) the whole set of variables is reorganised in the eLTER Controlled Lists in a hierarchical way with SO as the aggregation class of the variables they are member of.

Example SO:
Soil inventory (pedological/geological characterization) comprises following variables:

- soil texture
- soil hydraulic conductivity
- soil pH
- CEC
- BS
- bulk density
- base saturation
- soil type classification
- organic matter
- geological site characterization

All these variables will be organised as I-ADOPT variables. The SO will be represented as an I-ADOPT variableSet.