I-ADOPT Step-by-Step Guide

A presentation on how to apply the I-ADOPT Framework created by the GO FAIR Foundation in collaboration with the Interoperable Descriptions of Observable Property Terminology, in short, I-ADOPT Working Group of the Research Data Alliance, RDA.

Created by: GO FAIR Foundation

Version 1.0

Date: 3d of September 2024

I-ADOPT: Step-by-step guide for creating FAIR variable descriptions using the I-ADOPT Framework

RDA **I**nteroper**A**ble **D**escriptions of **O**bservable **P**roperty **T**erminology WG (I-ADOPT WG)

Core members:
Barbara Magagna, GO FAIR Foundation, NL
Gwenaëlle Moncoiffé, BODC, UK
Anusuryia Devaraju, CSIRO, AU
Maria Stoica, University of Colorado, US
Sirko Schindler, German Aerospace Center, DE
Alison Pamment, Centre for Environmental Data Analysis, UK

Created by

In collaboration with

This presentation helps to understand how you can create a FAIR variable descriptions using the I-ADOPT Framework step-by-step. This is for cases when the variables do not already exist in I-ADOPT compliant terminologies.

Links: https://www.rd-alliance.org/groups/interoperable-descriptions-observable-property-terminology-wg-i-adopt-wg/

Slide 2

Step-by-step guide

1. Identify components
2. Identify roles
3. Annotate with semantic concepts
4. Provide labels and description
5. Reuse existing or create an identifier reference

Providing a FAIR variable description based on the I-ADOPT Framework requires the following successive actions.

- Identify the components
- Identify their roles
- Annotate with concepts from FAIR terminologies
- Provide labels and description for the variable
- Enrich an existing variable concept with I-ADOPT modeling extension or create an identifier reference for the variable

---

1- Identify components

1.a Understand

A clear understanding of what your variable is, what phenomena it relates to or describes, is essential.

- What kind of values does it produce?
- Are they quantitative or qualitative observations?
- What units are usually associated with the variable?
- What methods are typically used to derive the values?

First be aware that a clear understanding of the phenomena that your variable should describe is essential. Either you have the domain expertise or you have access to domain experts to help with understanding how to decompose the variable into its essential components. In the analysis, try also to answer the following questions:

- What kind of values does it produce?
- Are they quantitative or qualitative observations?
- What units are usually associated with the variable?
- What methods are typically used to derive the values?

Slide 4

1- Identify components

1.b Analyse
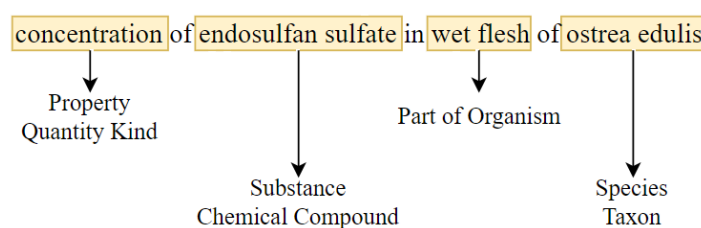
Identify the various components of the description.

concentration of endosulfan sulfate in wet flesh of ostrea edulis

Identify what components are needed to describe your observation. Components are reusable atomic parts of the variable description. Ask yourself the question: Could this component be a meaningful part of other variable descriptions as well? If the answer is yes, you most likely found a new component. If not, you probably need to split it up even more.

1- Identify components

1.c Generalize

The components are often only specialized variations of more generic concepts - identify what these could be:

concentration of endosulfan sulfate in wet flesh of ostrea edulis

- concentration → Property, Quantity Kind
- endosulfan sulfate → Substance, Chemical Compound
- wet flesh → Part of Organism
- ostrea edulis → Species, Taxon

The components are specialized variations of more generic concepts. Try to find what these more general concepts are. See the examples on the slide. Generalizing can help identify terminologies where you can find appropriate terms. Note that some of the terms may sometimes refer to elements of method, experimental condition, units of measurement. These should not be included in the variable description according to I-ADOPT but kept separate.

2 - Identify roles

2.a Look at associated data

Values and units can give important hints about the property and other components.

- For **qualitative** variables check the values - are they words, symbols? What do they describe? Are they from known controlled vocabularies?

- For **quantitative** variables check the units - what quantity kind(s) do they represent?

Values of "concentration of endosulfan sulfate in wet flesh of ostrea edulis" are measured, e.g., in micrograms per kilogram.

Values and units can give important hints for the further analysis.

- For qualitative variables check the kind of values - are they words, symbols? What do they describe? Are they from known controlled vocabularies?
- For quantitative variables check the units - what quantity kind(s) do they represent? Or are the values dimensionless?

The unit of measure or the vocabulary used for the values of a variable will help to narrow down the list of possible properties in the next step. For example if the variable has a derived unit of the form mass of something per mass of something else then this already gives a hint that a matrix concept will be needed.

Slide 7

2 - Identify roles

2.b Property

The Property is a generalized characteristic expressed by the value.

concentration of endosulfan sulfate in wet flesh of ostrea edulis
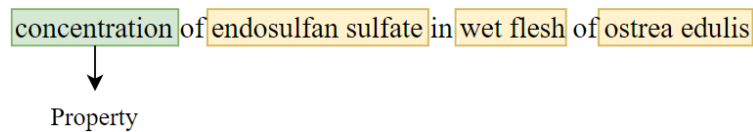
Property

You may use I-ADOPT's Unit-to-Property Lookup to deduce candidate properties from the unit

micrograw per kilogram

| Ontology ▲	Unit	Property
QUDT	Microgram Per Kilogram	• Mass Ratio

The Property is a generalized characteristic expressed by the value and is sometimes also called Quantity Kind. If the Property is not known, it may be deduced from the unit (e.g., using I-ADOPT Unit-to-Property Lookup) or from the vocabulary used to define possible values. In a more precise description we would specify which concentration we are looking for. Microgram per kilogram would lead to mass ratio or mass concentration.

Slides: https://i-adopt.github.io/terminologies/unit2property/

2 - Identify roles

2.c Object of Interest

The Object of Interest is the Entity whose Property is observed.

concentration of endosulfan sulfate in wet flesh of ostrea edulis

↓
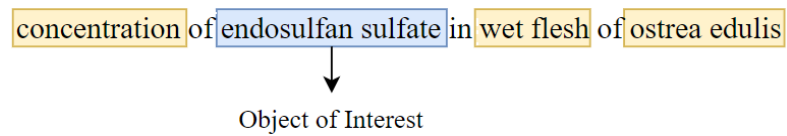
Object of Interest

The Object of Interest is one of the entities involved in the observation. In particular, it is the Entity whose Property is observed.

2 - Identify roles

2.d Matrix

The Matrix of the observation is entity in which the Object of Interest is embedded.

concentration of endosulfan sulfate in wet flesh of ostrea edulis

↓

Matrix

If the Object of Interest is embedded in or is a particular part of another Entity, that Entity is the Matrix of the observation. Note that not every observation necessarily contains a Matrix.

2 - Identify roles

2.e Context of Object(s)

Context Objects are all other Entities needed to describe the Variable.

concentration of endosulfan sulfate in wet flesh of ostrea edulis

Context Object

Context Objects are all other Entities needed to describe the Variable and provide essential contextual information to comprehensively describe the observation.

2 - Identify roles

2.f Further Decompose Entities *(if required)*

Check whether it is necessary to further decompose the identified entities into more general reusable concepts to constrain their scope in this particular scenario.

concentration of endosulfan sulfate in wet flesh of ostrea edulis

| Property | Object of Interest | Constraint | Matrix | Context object |
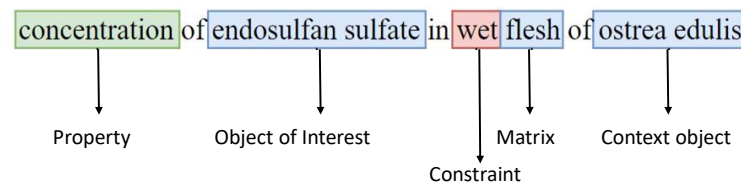
Revisit the identified Entities (Object of Interest, Matrix, and Context Object(s)) once again. Check whether it is possible or necessary to further decompose them into more general concepts and identify them as Constraints that confine their scope in this particular scenario.

3 - Annotate with semantic concepts

3.a Make your variable description machine readable

- Link each component to a concept from a commonly available terminology.
- Find suitable terminologies in the I-ADOPT Catalogue of Terminologies.
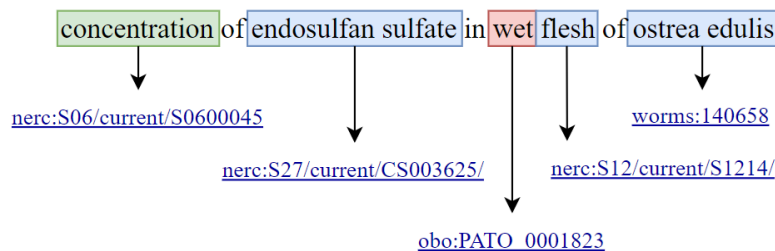
Make your variable description machine readable by linking each component to a **concept from a commonly available terminology**. You may consult the I-ADOPT Catalogue of Terminologies to find suitable terminologies. It lists terminologies that contains terms (or concepts) for the classes defined in I-ADOPT. Terminologies for Entities are not further distinguished as they may appear in different roles depending on the Variable.

Links: https://i-adopt.github.io/terminologies/

Property (concentration): http://vocab.nerc.ac.uk/collection/S06/current/S0600045

ObjectOfInterest (endosulfate sulfate):

http://vocab.nerc.ac.uk/collection/S27/current/CS003625

Matrix (flesh): http://vocab.nerc.ac.uk/collection/S12/current/S1214

ContextObject (Ostrea edulis):

http://marinespecies.org/aphia.php?p=taxdetails&id=140658

Constraint (for flesh) (wet): http://purl.obolibrary.org/obo/PATO_0001823

Slide 13

3 - Annotate with semantic concepts

| Earth Science Semantic Resources	List of Earth Science vocabulary repositories (includes BioPortal and many others)	http://bit.ly/EarthScienceSemanticResources
BioPortal	List of vocabulary resources from multiple domains (mostly biomedical); CEDAR can use these resources	https://bioportal.bioontology.org
AgroPortal	List of vocabulary resources related to agriculture; many are also in BioPortal	http://agroportal.lirmm.fr
EcoPortal	List of vocabulary resources related to ecology	https://ecoportal.lifewatch.eu/
BiodivPortal	List of vocabulary resources related to biodiversity	https://biodivportal.gfbio.org/
ESIP Community Ontology Repository (COR)	Ontologies related to earth science.	https://cor.esipfed.org
MMI Ontology Registry and Repository (ORR)	Ontologies related to marine science	https://mmisw.org/ont
Linked Open Vocabularies (LOV)	Somewhat idiosyncratic collection of RDF vocabularies on any topic	https://lov.linkeddata.es/dataset/lov/
Basic Register of Thesauri, Ontologies & Classifications	Collected information about vocabularies, terms, and terminology registries to facilitate use of knowledge organization systems.	https://bartoc.org
Linked Open Data Cloud	Source of all graphic images of Linked Open Data resources, this has very limited searching abilities	https://lod-cloud.net/

Another way to find appropriate concepts is to look into terminology catalogues like BioPortal. See here a useful list.

Earth Science Semantic Resources	http://bit.ly/EarthScienceSemanticResources
BioPortal	https://bioportal.bioontology.org
AgroPortal	http://agroportal.lirmm.fr
EcoPortal	https://ecoportal.lifewatch.eu/
BiodivPortal	https://biodivportal.gfbio.org/
ESIP Community Ontology Repository (COR)	https://cor.esipfed.org
MMI Ontology Registry and Repository (ORR)	https://mmisw.org/ont
Linked Open Vocabularies (LOV)	https://lov.linkeddata.es/dataset/lov/
Basic Register of Thesauri, Ontologies & Classifications	https://bartoc.org
Linked Open Data Cloud	https://lod-cloud.net/

4 - Provide labels and descriptions

4.a Label the Variable

- Variables can have two or more labels: the preferred label (mandatory) and alternative label(s) (optional)

- Labels should be unique, unambiguous and preferably consistent

- Labels can be constructed with components of the variable and follow a consistent grammar
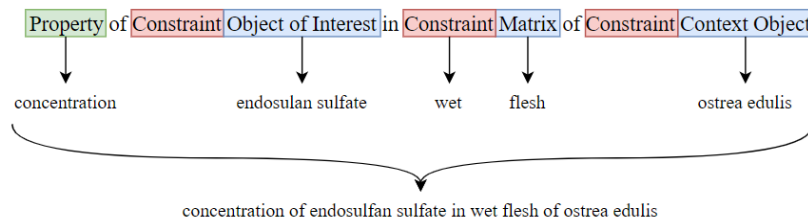
Think about how the variable needs to be labelled. Variables can have two or more labels - a preferred one which is mandatory and optional alternative ones. Labels should be unique, preferably consistent and unambiguous. The preferred label can be one used in your community, but make sure that you and the community fully agree on the meaning (see next slide). Labels can be constructed with components of the variable and follow a consistent grammar.

4 - Provide labels and descriptions

4.b Add a definition

- Provide a concise human-readable text defining the variable

- If necessary, include permanent links to online material with additional contextual information

- The aim of the description is to help humans better understand the variable, its applications and specificity

Any concept in a terminology should have a description. So do variables. A description or definition helps humans understand better the variable.
Add links to an online reference if it provides contextual information on how the variable is measured like the method and protocol applied.

5 - Reuse or create an identifier reference

5.a Enrich an existing Variable concept (if applicable)

Check whether you can reuse an existing variable and attach the identified components using I-ADOPT references.

If a variable concept with an identifier already exists, you can reuse it and enrich it with I-ADOPT associations. (iop:hasProperty, iop:hasObjectOfInterest, iop:hasMatrix, iop:hasContextObject, iop:hasConstraint).

Link: https://i-adopt.github.io/ontology/

5 - Reuse or create an identifier reference

5.b Create a Variable concept

Provide an identifier reference using Linked Data Principles and attach the identified components using I-ADOPT references.

If you can't reuse an existing variable, get a URI for the Variable for its identifier reference using Linked Data Principles (like SKOS) and attach the previously identified components according to the I-ADOPT ontology (iop:hasProperty, iop:hasObjectOfInterest, iop:hasMatrix, iop:hasContextObject, iop:hasConstraint).

Link: https://i-adopt.github.io/ontology/

Slide 18

A FAIR variable representation in RDF

Example in turtle (excerpt, see full definition here):

```
@prefix nercP01: <http://vocab.nerc.ac.uk/collection/P01/current/> .
@prefix nercS06: <http://vocab.nerc.ac.uk/collection/S06/current/> .
@prefix nercS12: <http://vocab.nerc.ac.uk/collection/S12/current/> .
@prefix nercS27: <http://vocab.nerc.ac.uk/collection/S27/current/> .
@prefix iadopt: <https://w3id.org/iadopt/ont/> .
@prefix rdfs:   <http://www.w3.org/2000/01/rdf-schema#> .
@prefix worms:  <http://marinespecies.org/aphia.php?p=taxdetails&id=> .

…
  nercP01:IC000344
  a          iadopt:Variable ;
  rdfs:label "concentration of endosulfane sulfate in wet flesh of ostrea edulis"@en ;
  iadopt:hasObjectOfInterest  nercS27:CS003625 ;
  iadopt:hasProperty          nercS06:S0600045 ;
  iadopt:hasMatrix            nercS12:S1214 ;
  iadopt:hasContextObject     worms:140658 ;
  iadopt:hasConstraint [
    a          iadopt:Constraint ;
    rdfs:label "wet"@en ;
    iadopt:constrains nercS12:S1214 ;
  ] ;
  .
```

Once published as a semantic concept according to the I-ADOPT Framework it can be downloaded in RDF as a rich knowledge graph.

Link: https://raw.githubusercontent.com/i-adopt/examples/main/templates/full.ttl

Acknowledgements

Created by: GO FAIR Foundation   **In collaboration with:** RDA RESEARCH DATA ALLIANCE I-ADOPT WG