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

RESEARCH DATA ALLIANCE
I-ADOPT WG

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

GO FAIR foundation
CC BY NC ND

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?

GO FAIR foundation
CC BY NC ND

1- Identify components

1.b Analyse

Identify the various components of the description.

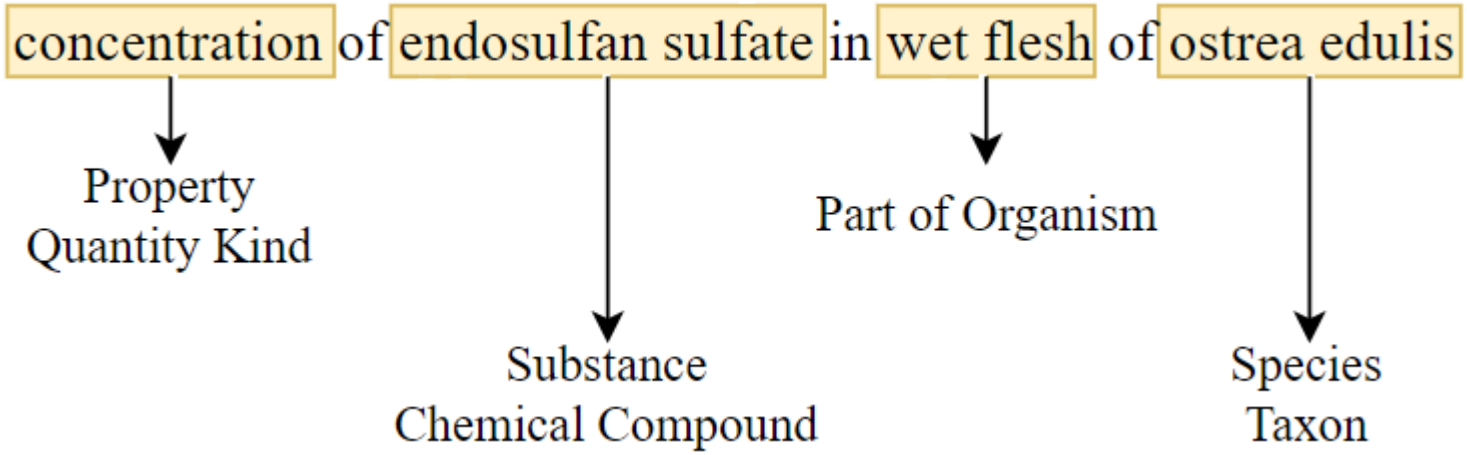
concentration of endosulfan sulfate in wet flesh of ostrea edulis

GO FAIR foundation
CC BY NC ND

1- Identify components

1.c Generalize

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



GO FAIR foundation
CC BY NC ND

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?

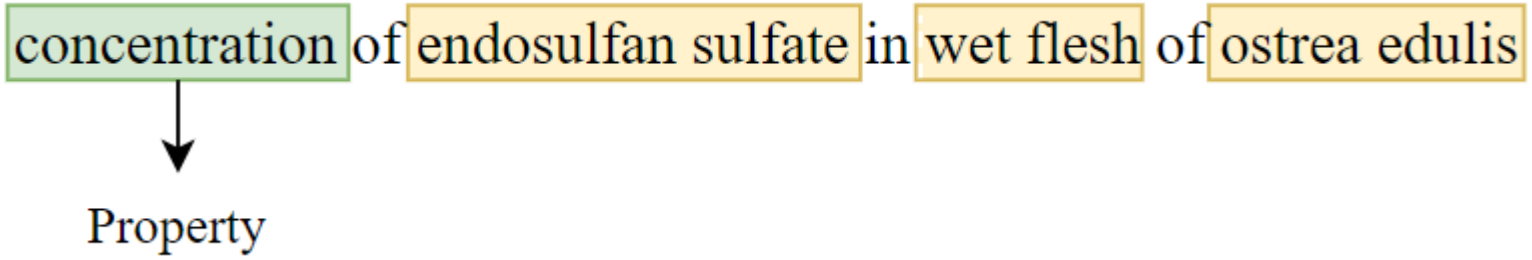
Values of “concentration of endosulfan sulfate in wet flesh of ostrea edulis” are measured, e.g., in micrograms per kilogram.

GO FAIR foundation
CC BY NC ND

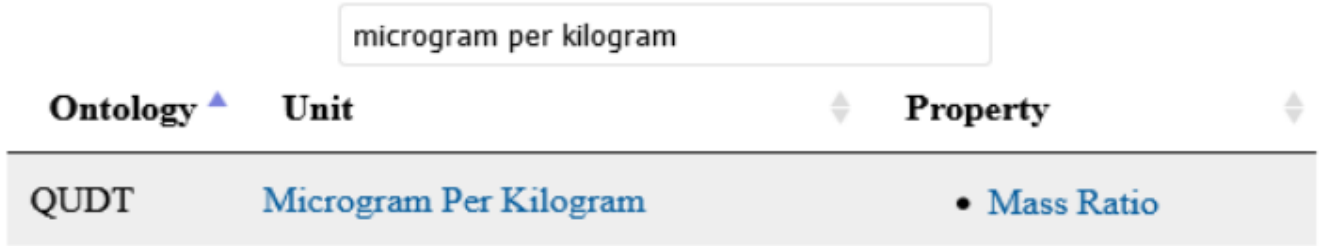
2 - Identify roles

2.b Property

The Property is a generalized characteristic expressed by the value.



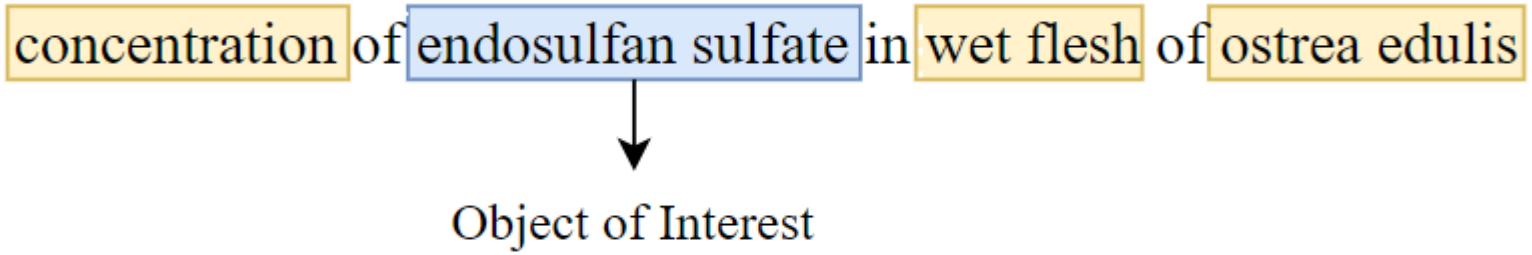
You may use I-ADOPT's Unit-to-Property Lookup to deduce candidate properties from the unit



2 - Identify roles

2.c Object of Interest

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

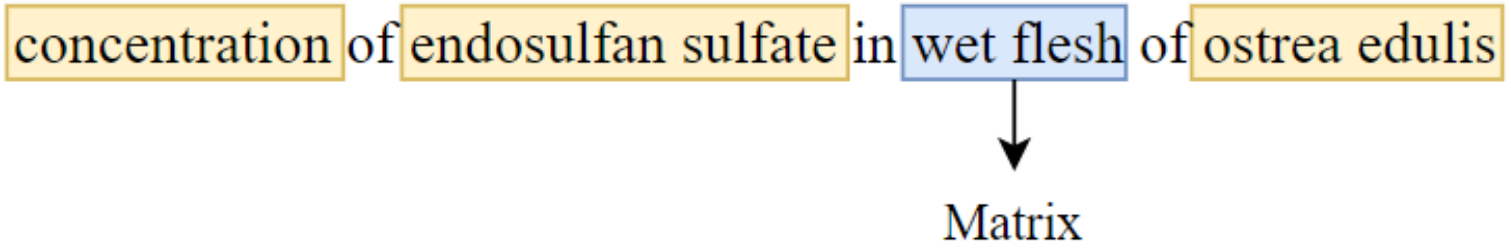


GO FAIR foundation
CC BY NC ND

2 - Identify roles

2.d Matrix

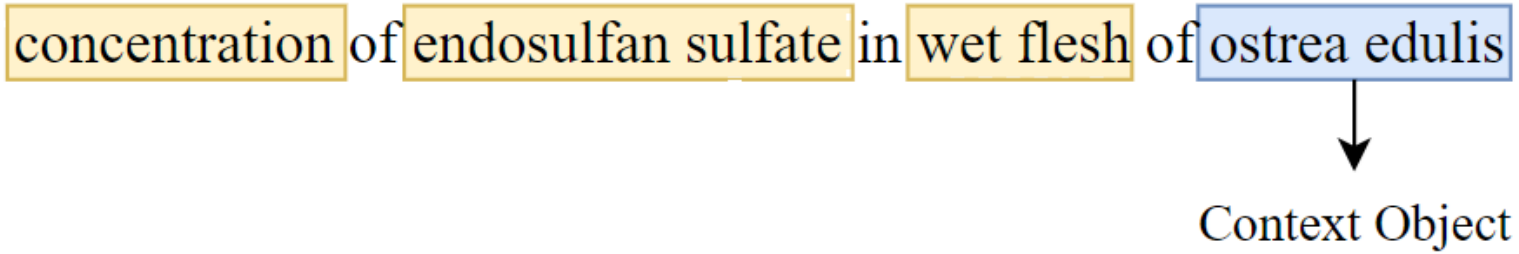
The Matrix of the observation is entity in which the Object of Interest is embedded.



2 - Identify roles

2.e Context of Object(s)

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

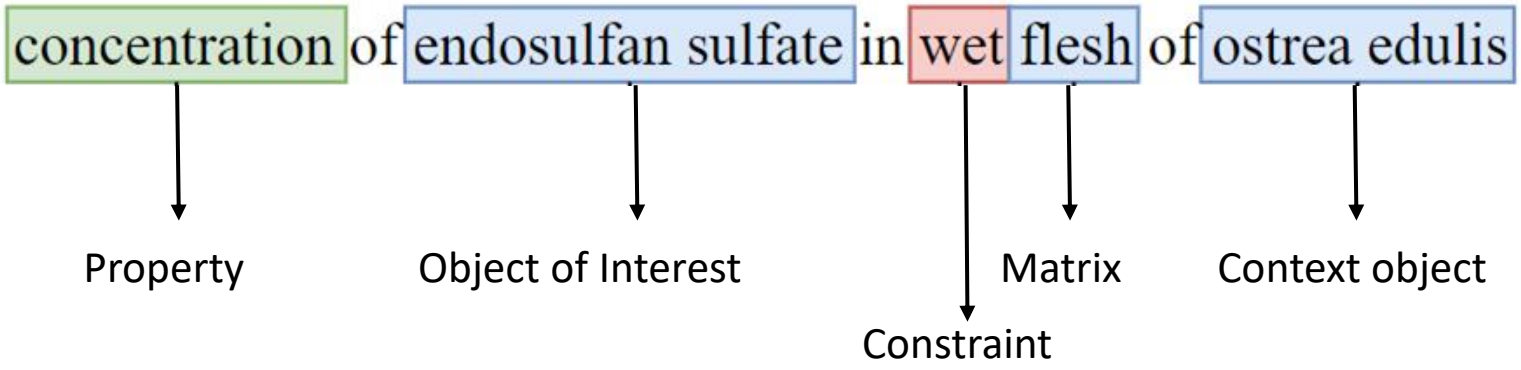


GO FAIR foundation
CC BY NC ND

2 - Identify roles

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

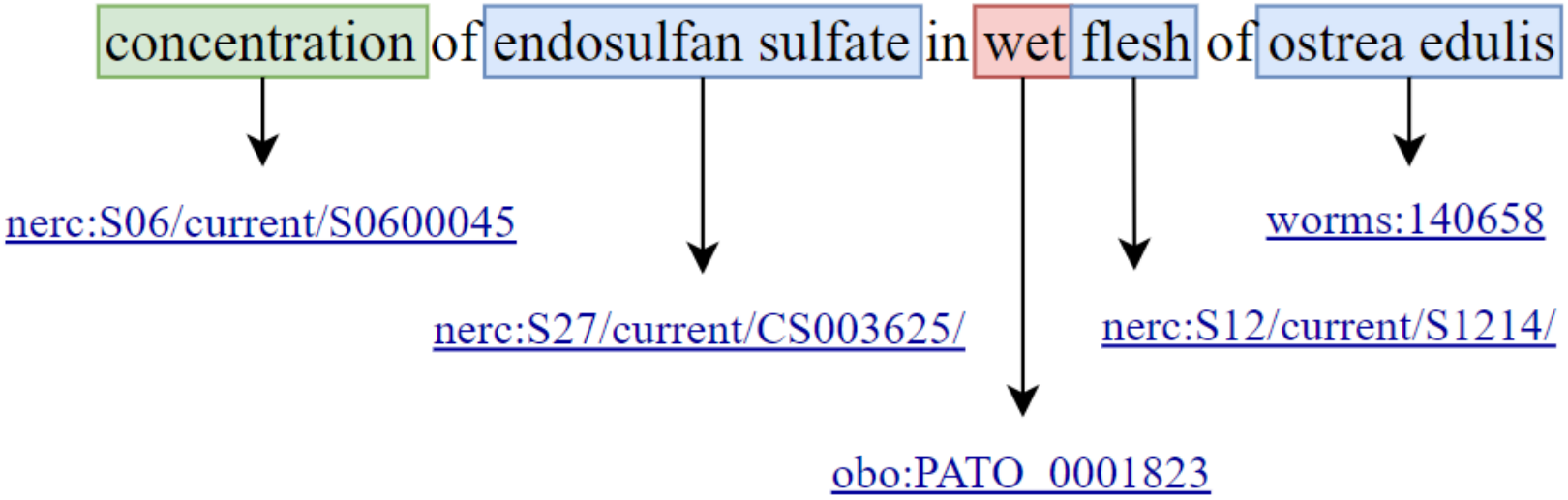
Check whether it is necessary to further decompose the identified entities into more general reusable concepts to constrain their scope in this particular scenario.



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.



GO FAIR foundation
CC BY NC ND

3 - Annotate with semantic concepts

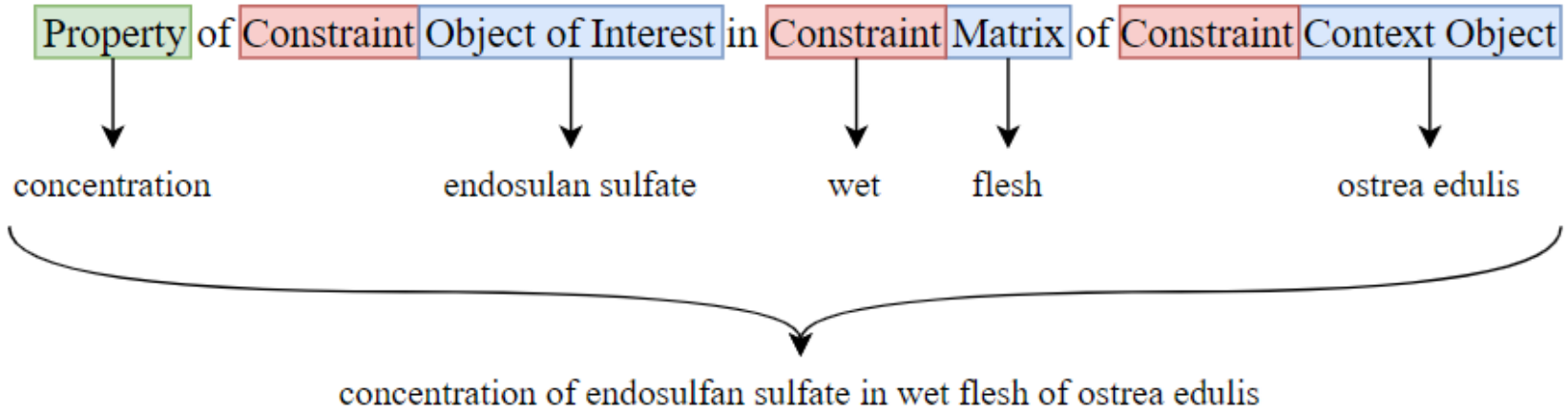
Earth Science Semantic Resources	List of Earth Science vocabulary repositories (includes BioPortal and many others)	http://bit.ly/EarthScienceSemanticResources
BioPortal	List of vocabulary resources from multiple domains (mostly biomedical); CEDAR can use these resources	https://bioportal.bioontology.org
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/

GO FAIR foundation
CC BY NC ND

4 - Provide labels and descriptions

4.a Label the Variable

- Variables can have two labels: the preferred label (mandatory) and the alternative label (optional)
- Labels should be unique, unambiguous and preferably consistent
- Labels can be constructed with components of the variable and follow a consistent grammar



GO FAIR foundation
CC BY NC ND

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

GO FAIR foundation
CC BY NC ND

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.

GO FAIR foundation
CC BY NC ND

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.

GO FAIR foundation
CC BY NC ND

A FAIR variable representation in RDF

Example in turtle (excerpt, see full description 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 ;
  ] ;
  .
```

Acknowledgements and further reading

License for this presentation: CC BY-NC-ND

- Reusers may copy and distribute the material in any medium or format in unadapted form only, for noncommercial purposes only, and only so long as attribution is given to the creator.

Created by: GO FAIR Foundation

In collaboration with: RDA RESEARCH DATA ALLIANCE I-ADOPT WG

GO FAIR Foundation BY NC ND