



RDA WG Interoperability of Observable Property Description

I-ADOPT WG

**Barbara Magagna, Gwenaelle Moncoiffé, Anusuriya Devaraju, Maria Stoica,
Sirko Schindler, Alison Pamment**

Barbara Magagna

I-ADOPT WG - Overview

Motivation: Addressing the “I” of FAIR Data Management

By building a conceptual framework to support interoperability between existing Terminologies for variables and address a broad range of known user requirements

By promoting the use of FAIR terminologies to annotate research data with well identified, unambiguous and machine readable variable descriptions

I-ADOPT chairs: Barbara Magagna, Gwenaelle Moncoiffé, Maria Stoica, Anusuriya Devaraju

RDA WG I-ADOPT Wiki

- The WG already organized five official **RDA events**:
- 11 RDA Plenary meeting in Berlin: [Task group formed under Vocabulary Semantic Service Interest Group](#) (VSSIG)
- 13 RDA Plenary meeting in Philadelphia: [BOF- Harmonising FAIR descriptions of observational data](#)
- 14 RDA Plenary meeting in Helsinki: [Kick-Off Meeting of I-ADOPT](#)
- 15 RDA Virtual Plenary meeting: [Working Session Minutes, recording](#)
- 16 RDA Virtual Plenary meeting: [Working Session folder](#)

RDA WG I-ADOPT



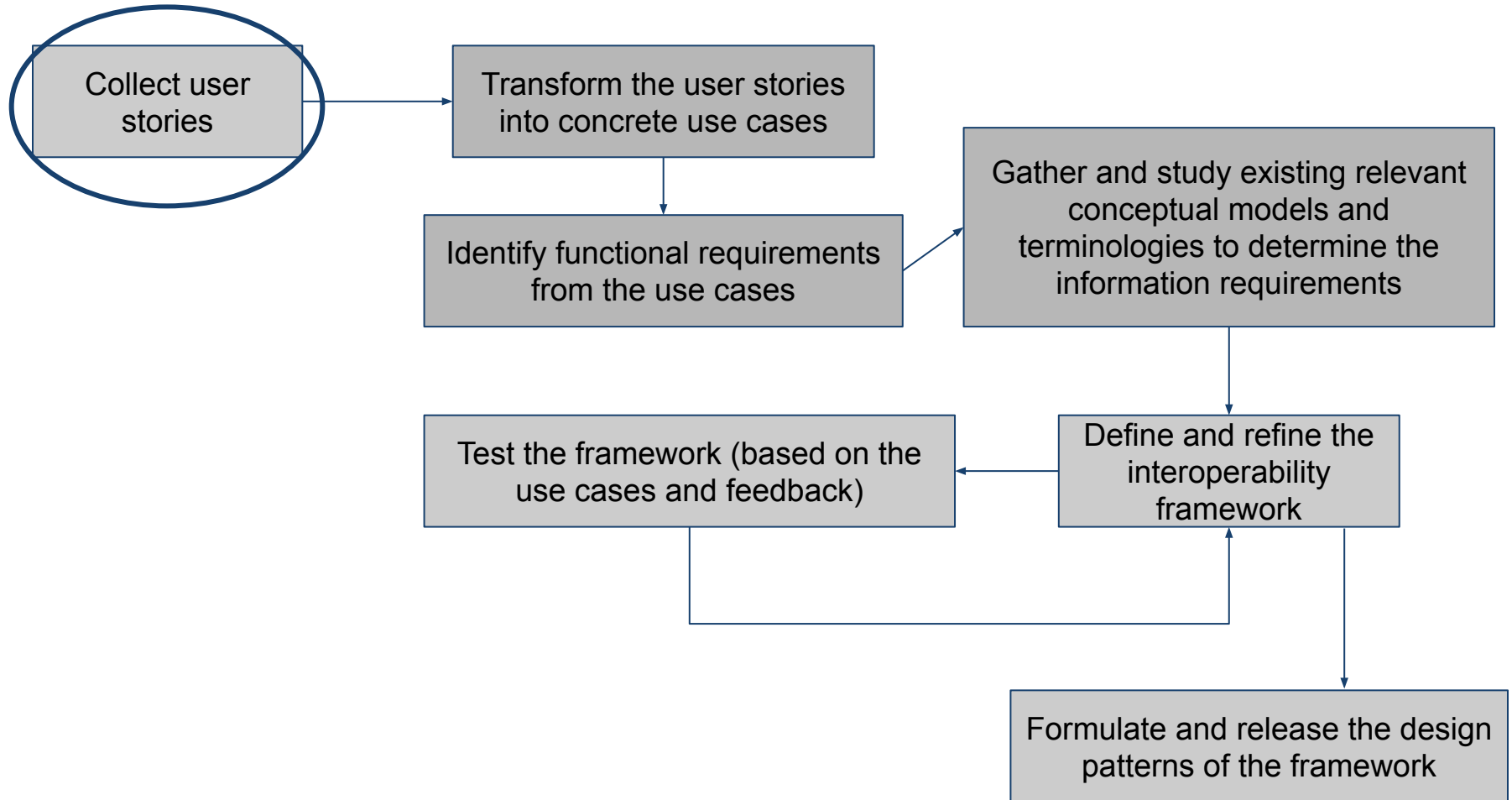
Task 1: Collect user stories and formalise into use cases	Nov 19 - February 20
Task 2: Survey observation-centric terminologies	Jan 20 - February 20
Task 3: Derive use case requirements	March 20 - April 20
Task 4: Analyse semantic representation of variables against requirements	May 20 - December 20
Task 5: Develop Interoperability Framework	Jan 21- April 21
Task 6: Test local mapping design patterns	May 21 - June 21

More details to be found in the [case statement](#)

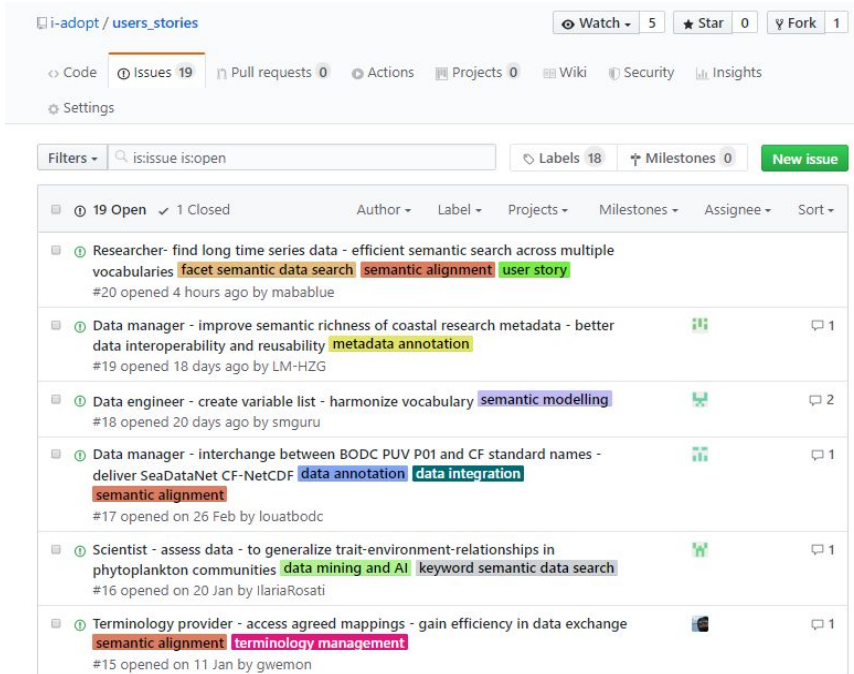
Maria Stoica/Anusuriya Devaraju

I-ADOPT - Methodology

High-Level Workflow



I-ADOPT user stories



i-adopt / users_stories

Watch 5 Star 0 Fork 1

Code Issues 19 Pull requests 0 Actions Projects 0 Wiki Security Insights

Settings

Filters is:issue is:open Labels 18 Milestones 0 New issue

19 Open	1 Closed	Author	Label	Projects	Milestones	Assignee	Sort
Researcher - find long time series data - efficient semantic search across multiple vocabularies	facet semantic data search	semantic alignment	user story	#20	opened 4 hours ago	by mabablue	
Data manager - improve semantic richness of coastal research metadata - better data interoperability and reusability	metadata annotation	#19	opened 18 days ago	by LM-HZG		1	
Data engineer - create variable list - harmonize vocabulary	semantic modelling	#18	opened 20 days ago	by smguru		2	
Data manager - interchange between BODC PUV P01 and CF standard names - deliver SeaDataNet CF-NetCDF	data annotation	data integration	semantic alignment	#17	opened on 26 Feb	by louatbodc	1
Scientist - assess data - to generalize trait-environment-relationships in phytoplankton communities	data mining and AI	keyword semantic data search	#16	opened on 20 Jan	by IlariaRosati	1	
Terminology provider - access agreed mappings - gain efficiency in data exchange	semantic alignment	terminology management	#15	opened on 11 Jan	by gwemon	1	

As a...

Data manager

I want to...

to use both BODC PUV P01 and CF standard name terms interchangeably

So that...

So that our Data Assembly Centre can deliver fully CF compliant formats such as SeaDataNet CF-NetCDF files (which require both P01 and CF names) and easily deliver and transform between multiple formats such as EGO and Ocean Data View.

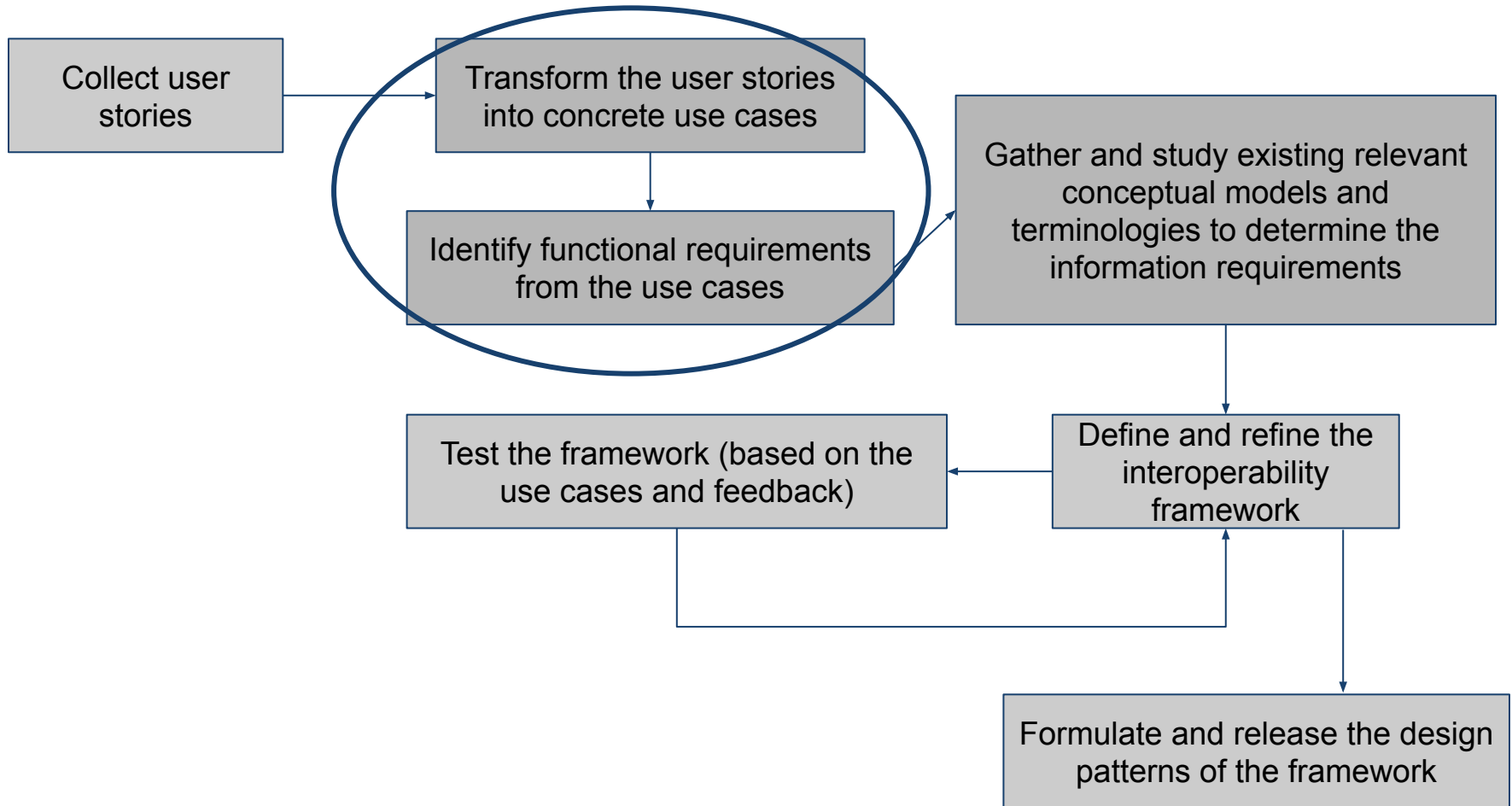
Domain(s)

Marine and atmospheric

Addition Information

SeaDataNet transport formats - <http://doi.org/10.13155/56547>

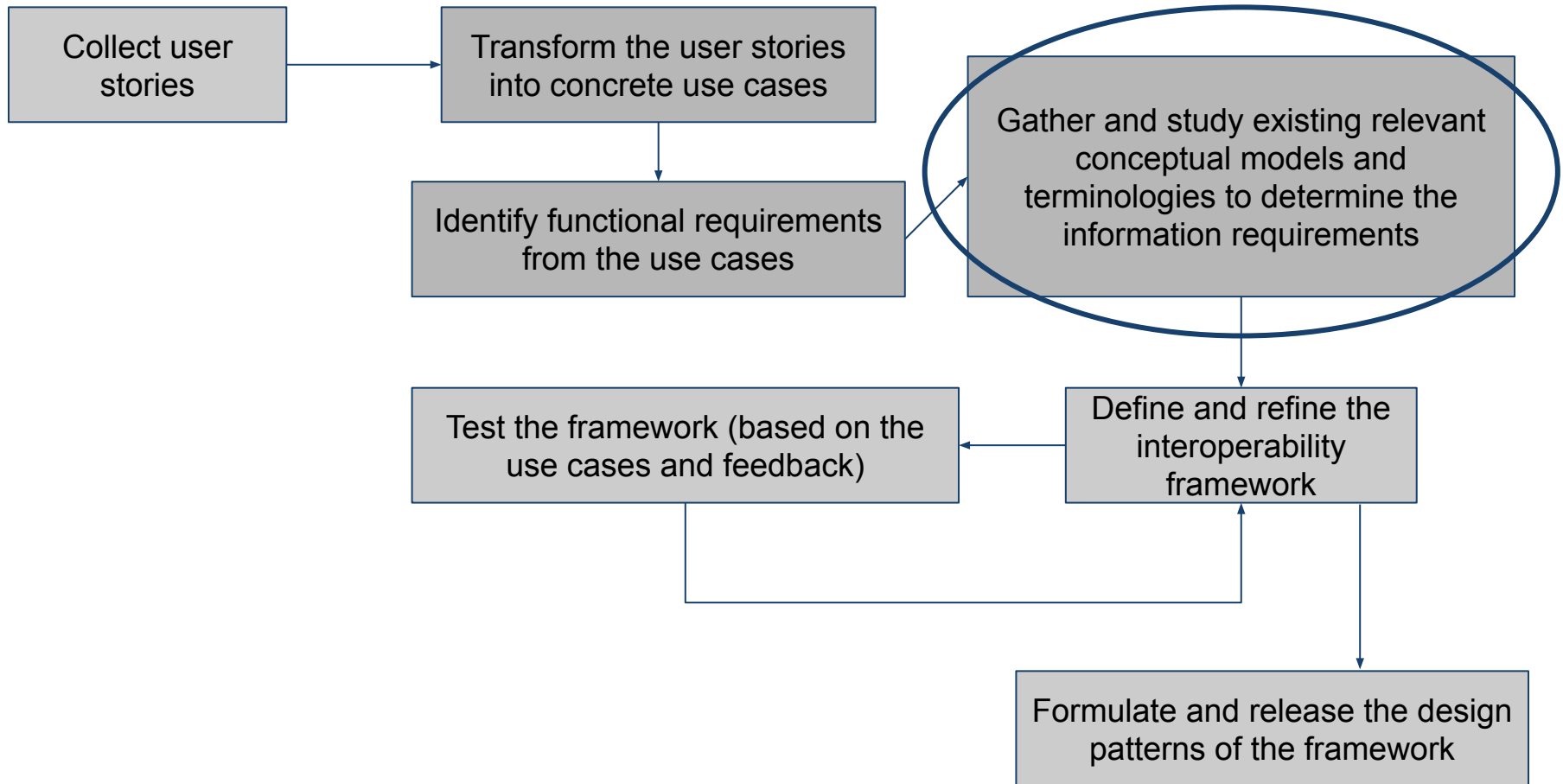
High-Level Workflow



Functional Requirements

Use case group	Brief Description	Framework requirements
Terminology generation and management	Gather and organize terminologies (inc synonyms and multi-language), encode them semantically, implement searching algorithms	Components are able to categorize/cover the full terminology landscape present across the relevant terminologies
Resource annotation	Automated or semi-automated annotation tools and services for resources	Components are flexible enough to be applied to the different forms and representations of the relevant terminologies
Data discovery and exploration	Keyword and faceted semantic data search	Enable terminology mapping through synonyms/altlabels and meaningful component categorizations
Data manipulation and integration	Combine and transform datasets from various sources and from different presentations/formats	Use only a limited (minimum) number of highly generic, essential components that apply across applications & terminologies

High-Level Workflow



Information Requirements : Terminologies Analysis

Catalogue of conceptual models and terminologies

B	C	D	E	H	I	J	K	L	M	N	O	P	Q	R	S	T
Reviewed by	Valid?	terminology name	access URL	type of terminology	domain-specificity	main language	other language(s)	Units and dimensions	Terms for observable property	Terms for research object	terms for constraints	terms for matrix	terms for context	Supported by a model framework?	Terms for method description?	Terms for attribute values?
Gwen	yes	French Crop Usage Thesaurus	http://ontology.irstea.fr/cropusage	thesaurus	narrow	FR		no	no	yes	no	no	yes	no	no	no?
Gwen	yes	Biodiversity Thesaurus	https://www.lotere.fr/skosmos/BLH/en	thesaurus	broad	EN	FR	no	yes	yes	no	yes	yes	no	no	no?
Gwen	yes	Wheat Trait Ontology	http://agroportal.lirmm.fr/ontologies/WHEATPHENOTYPE	thesaurus	narrow	EN		no	yes	no?	yes	no	no	no	no	yes
	yes	Minimal Information about any Sequence Ontology (MlxS) cryosphere extension	https://biportal.bioontology.org/ontologies/MIXS		narrow			no						no		
Gwen	yes	Complex Properties Model - as related to ocean measurements		ontology/model framework	generic	EN		no	no	no				yes	no	no

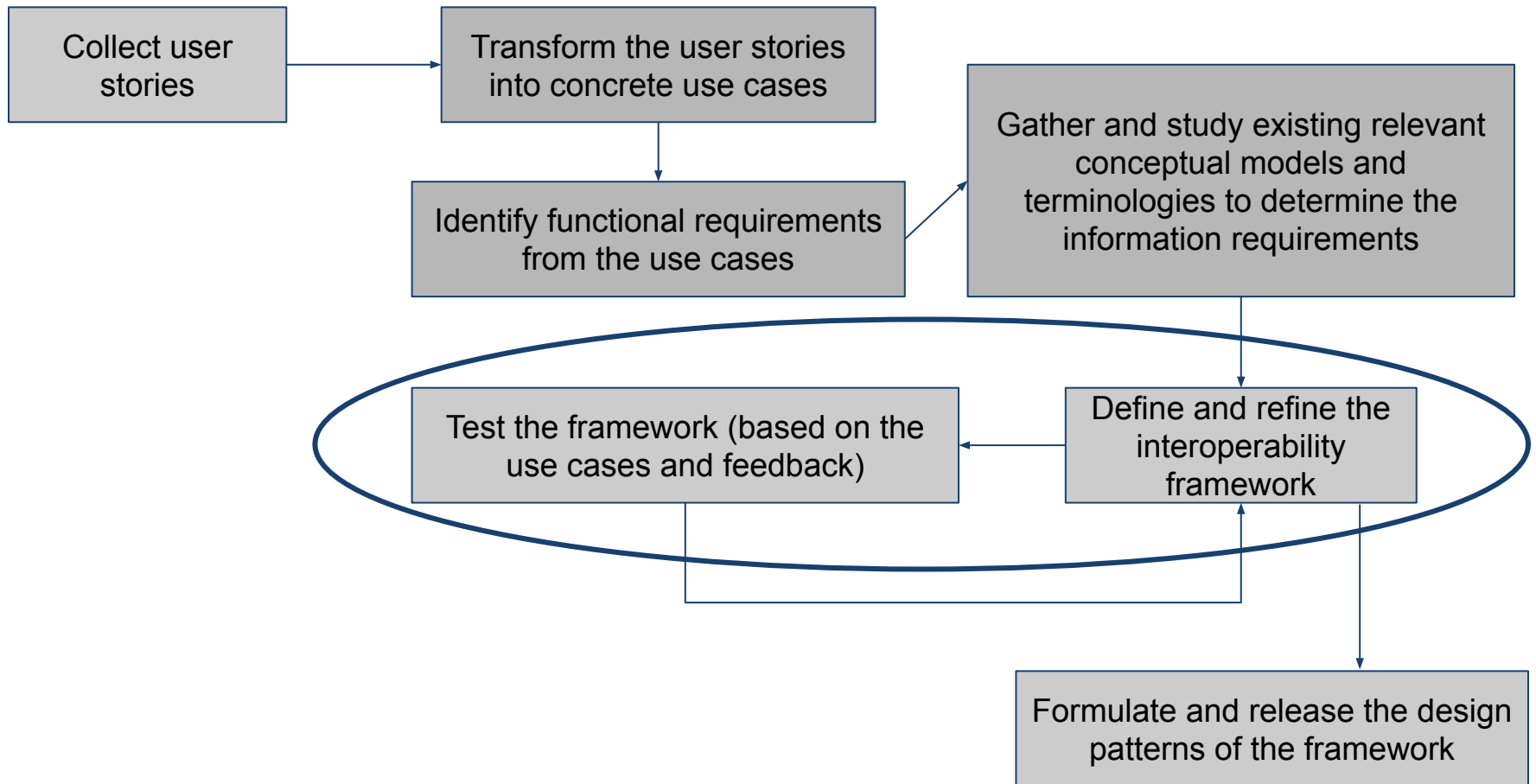
Terminologies Analysis

- Catalogued around 100 conceptual models and terminologies
- Found the following components occurred frequently:
 - objects (45)
 - properties/quantities (42)
 - matrix/medium (9+)
 - constraint (11+)
 - methodology (11)
- Although terms are present for these components, they are not always decomposed and atomized at this level (i.e., sometimes these components are found in combination)

Calanus finmarchicus, female, prosome length [mm]

Short name: C. finmarchicus f pros l [mm]

High-Level Workflow



Barbara Magagna

The I-ADOPT Framework Objectives

I-ADOPT Framework – Focus and Scope

Variable names are used everywhere but they are often quite fuzzy in their description!

A variable can be represented by a combination of descriptive components giving meaning to the value derived from a data acquisition event be it an observation, a measurement, a simulation or a calculation.

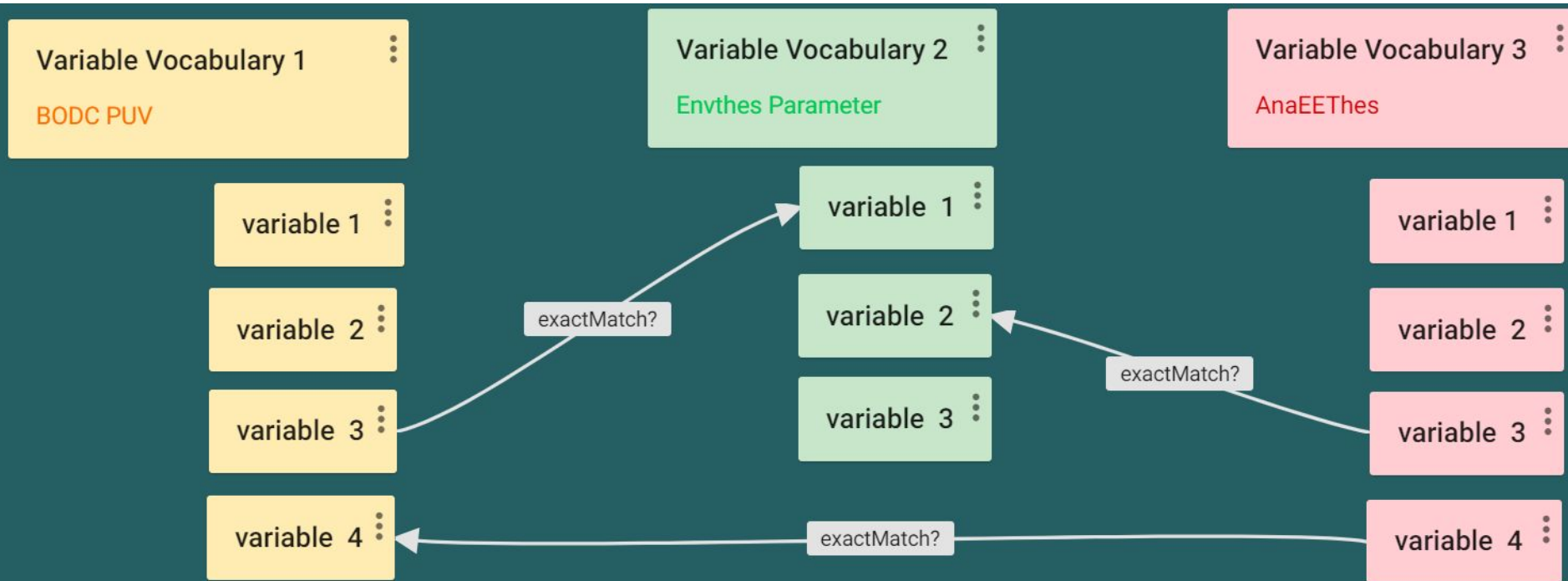
The concept “Variable” in I-ADOPT represents **WHAT** has been observed independently of WHERE, HOW and WHEN the data acquisition took place.

I-ADOPT only supports the **classification of types**.

.

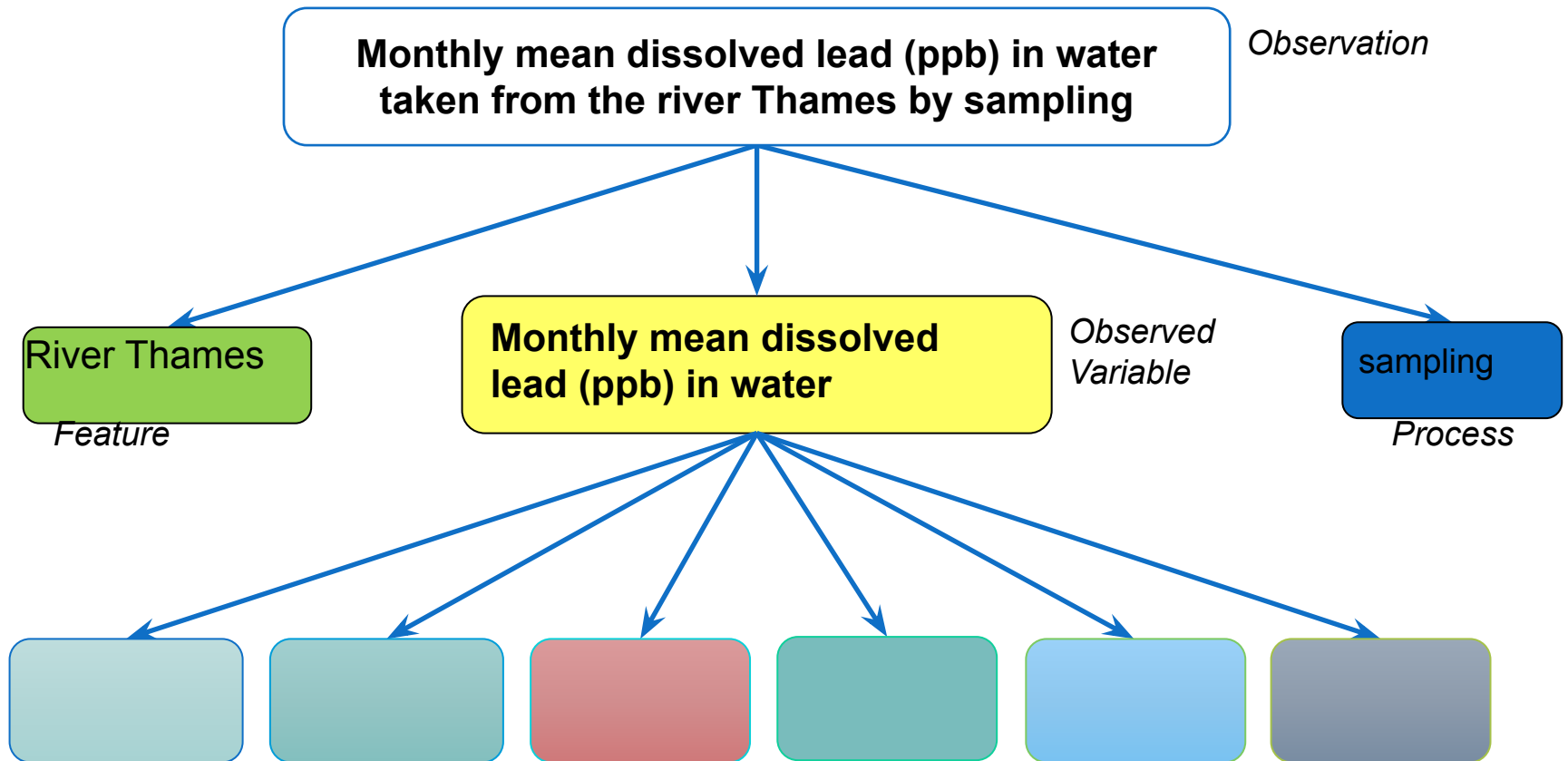
Addressed Challenge:

Proliferation of unmapped variable description terminologies

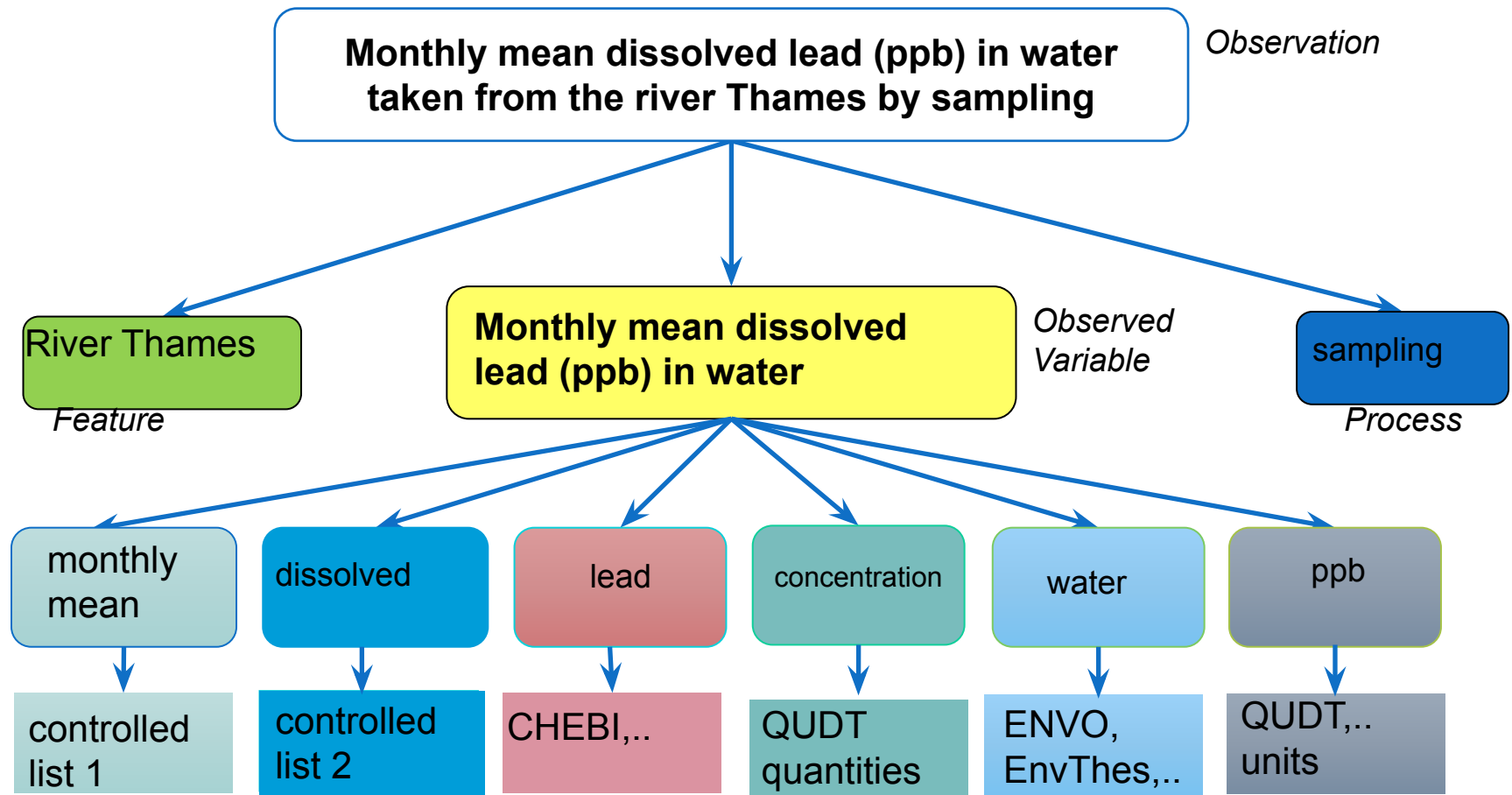


Examples: wind speed (vs speed of wind), soil colour (vs colour of soils). molar concentration of atomic nitrogen in earth's atmosphere (vs. nitrogen concentration)

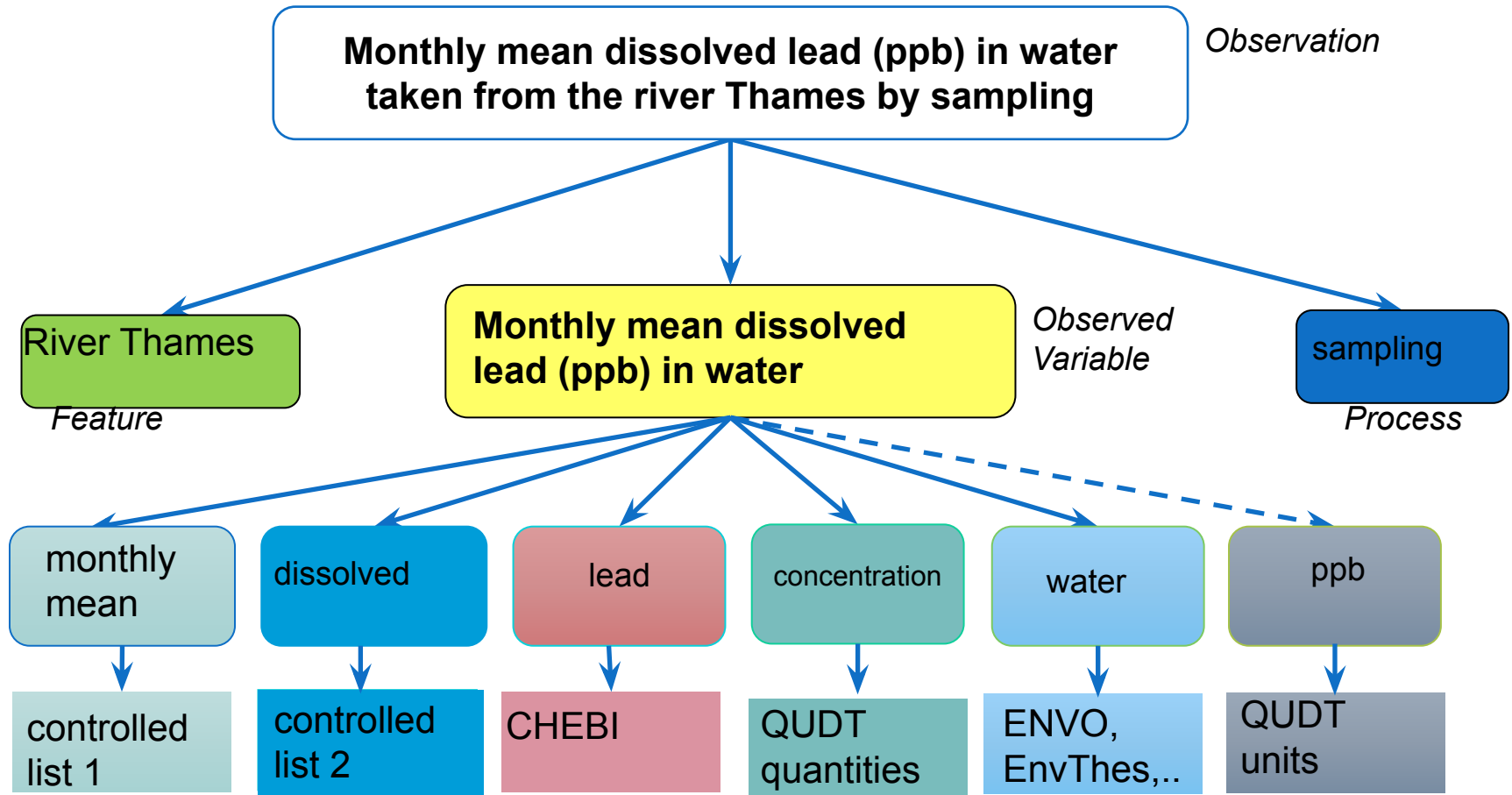
Conceptualizations of variable descriptions



Controlled vocabularies for atomic parts of variable descriptions



Controlled vocabularies for atomic parts of variable descriptions



I-ADOPT Framework – Overall objective and approach

Support interoperability between existing terminologies by

- ❑ Enabling semantically precise and FAIR descriptions of observed/measured/modelled variables used primarily in environmental research (like biodiversity, climatology)
- ❑ Decomposing descriptions into atomic components that are easier to map and reuse
- ❑ Developing the framework iteratively: its components are refined by reviewing current annotation practices and by validating against a set of requirements from real-world use cases with feedback from WG members
- ❑ Using examples where each component will be linked to concepts from existing vocabularies making these descriptions of observed variables machine readable
- ❑ Testing the I-ADOPT framework for interoperability between a number of existing terminologies.

I-ADOPT Framework – Outlook

- ❑ We recommend managers of SKOS vocabularies for labelling variables to use skos extended relations to map to the components of the framework to help align across vocabularies.
- ❑ We provide best practices to address different use cases and varying degrees of complexity of variables.
- ❑ In an associated effort, WG members will formalize the framework into an ontology.
- ❑ Context-sensitive, user friendly data annotation tools may then use this ontology to suggest appropriate FAIR terminologies for variable descriptions.

A. Identifying essential components needed to describe FAIR Variables

Examples of
components
(not exhaustive):

I-ADOPT Framework

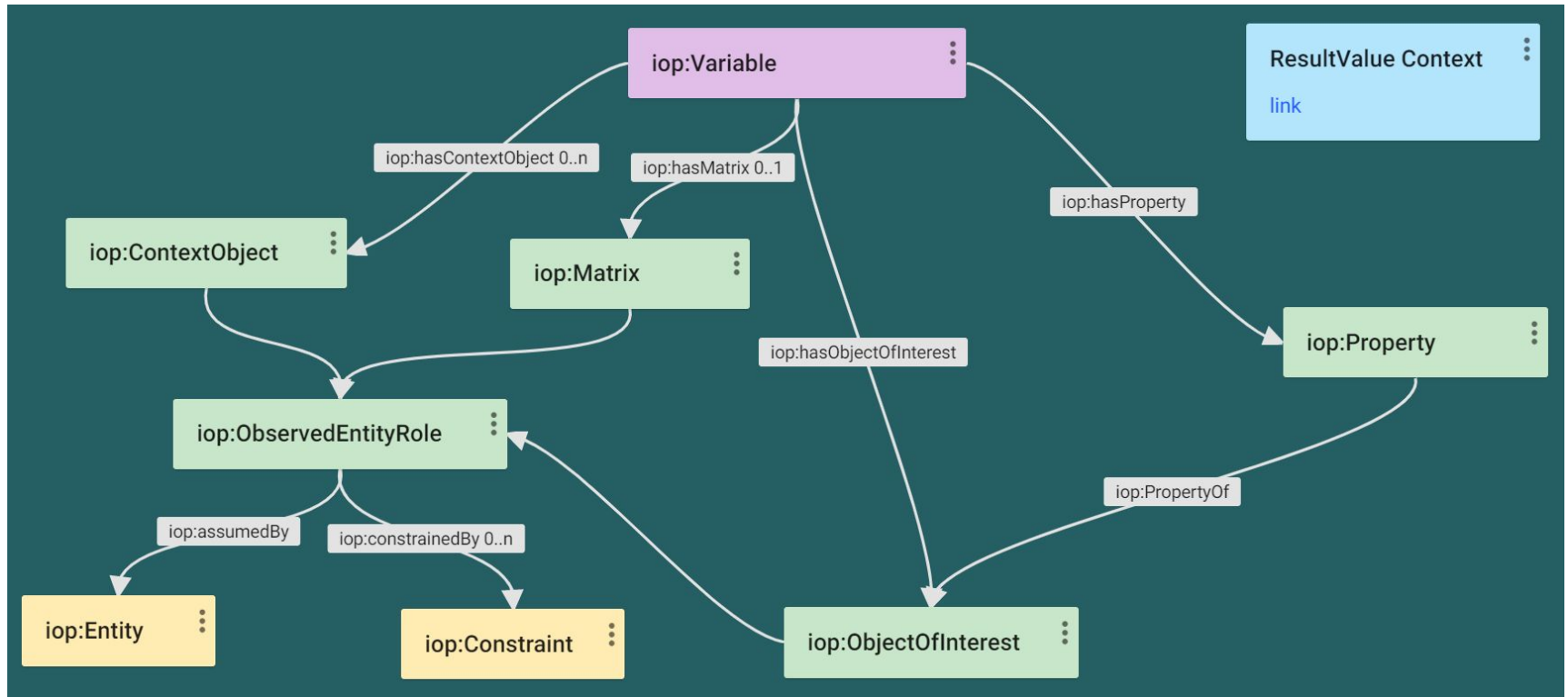
IOP Object of Interest

IOP Property

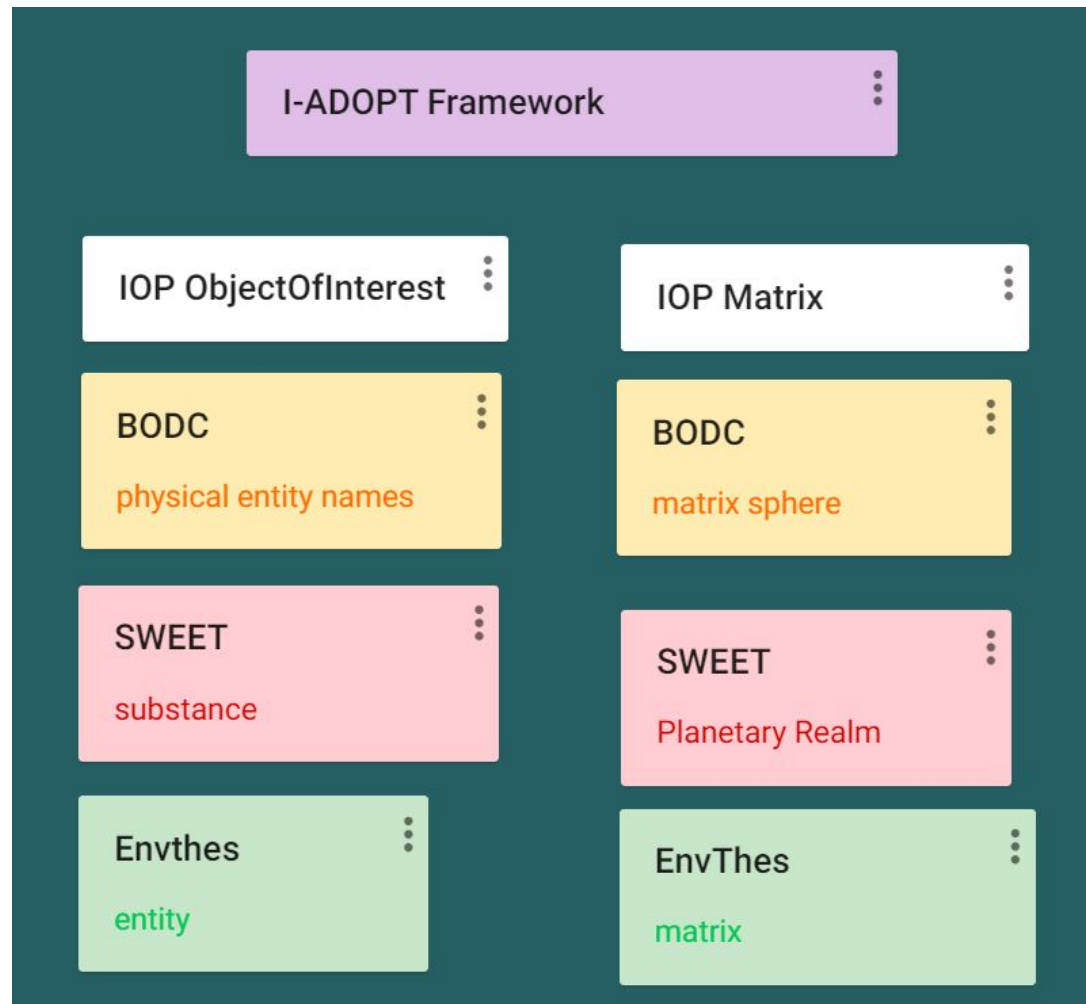
IOP Matrix

IOP Constraint

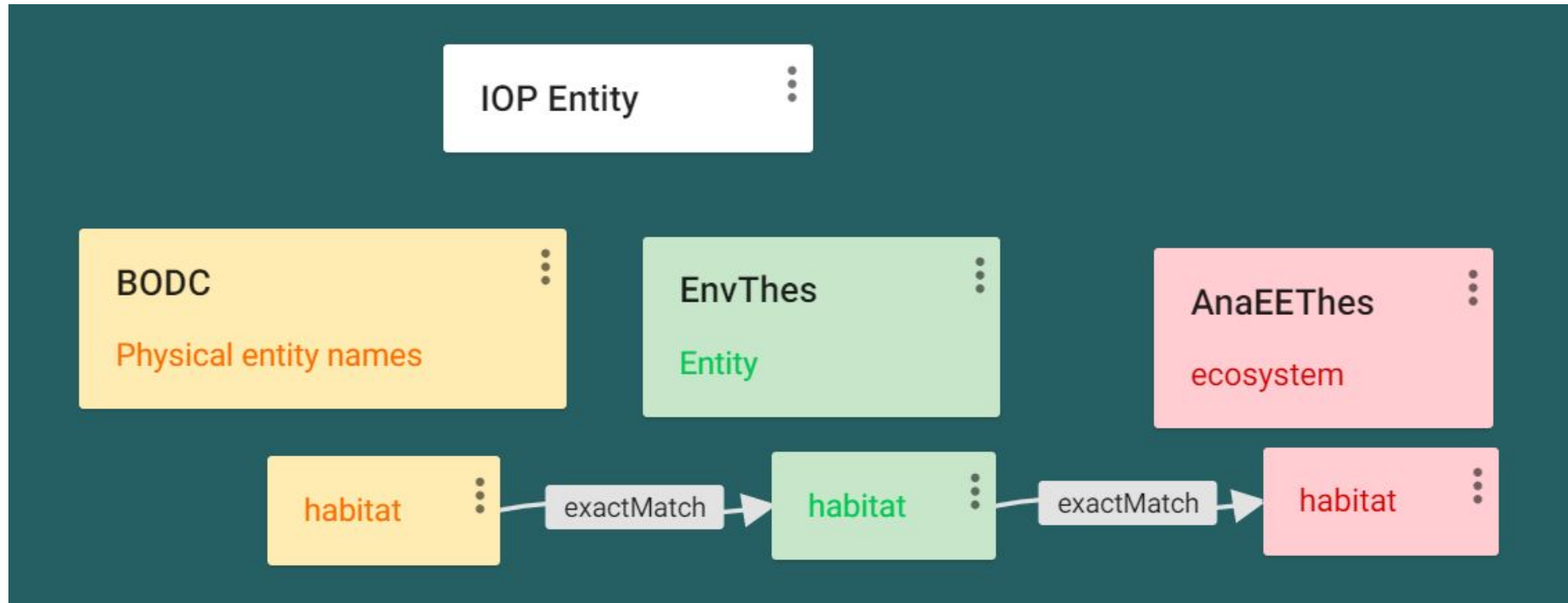
B. Identifying relations between the components -> I-ADOPT Framework



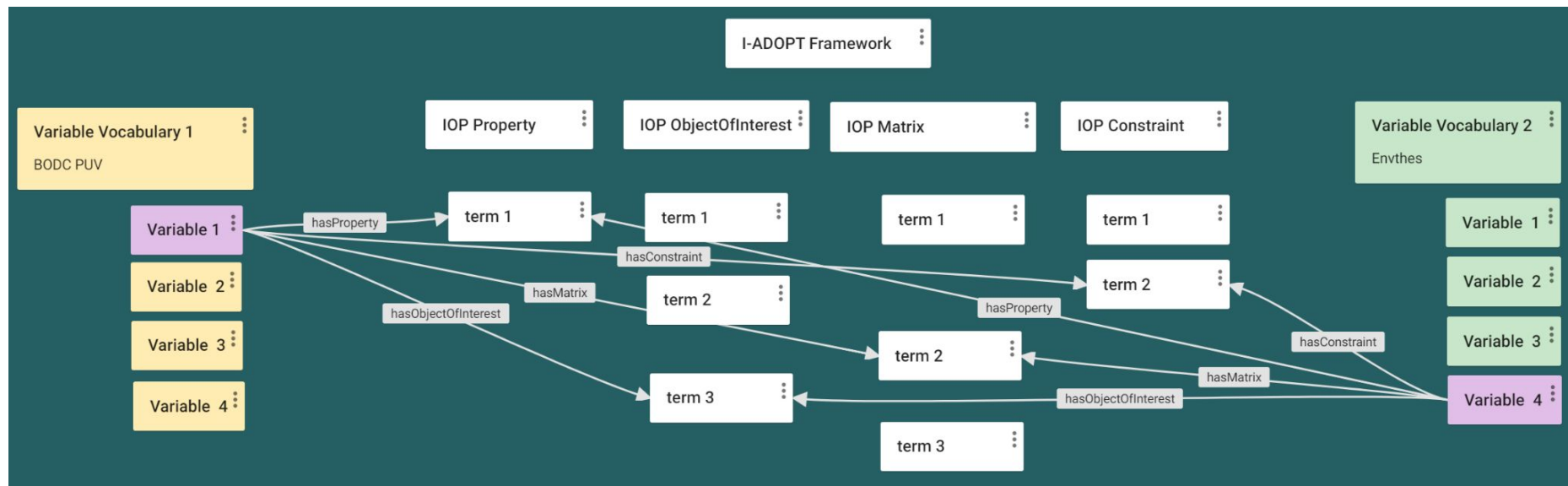
C. Identifying suitable terminologies for the framework components



D. Mapping terms between framework components terminologies

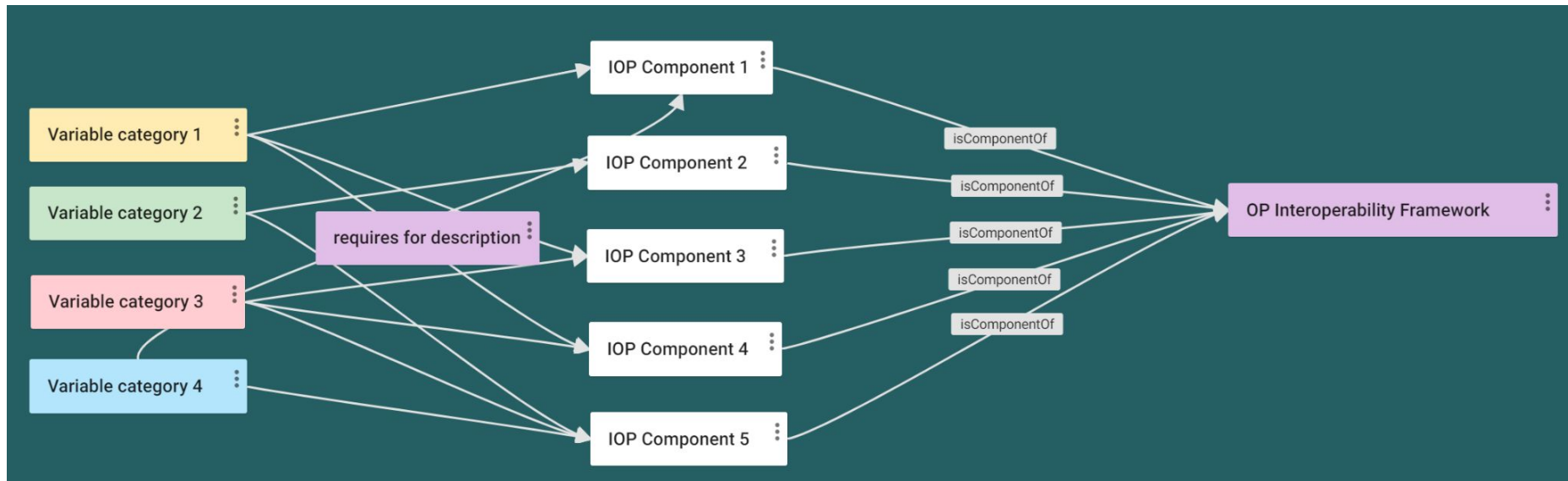


E. Providing means for identifying the same variables for mappings across vocabularies

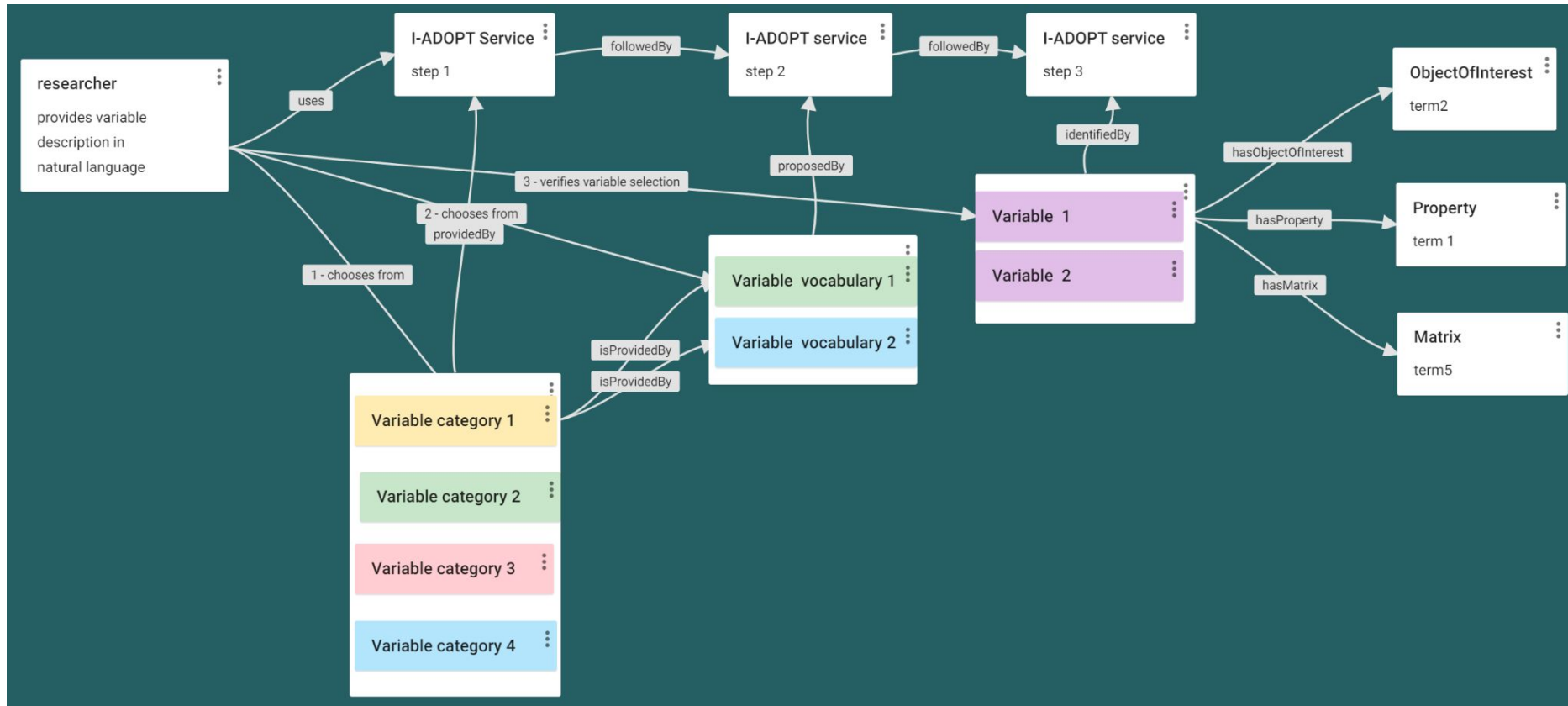


F. Providing guidelines how to represent accurately new variables based on identified categories and description patterns

:

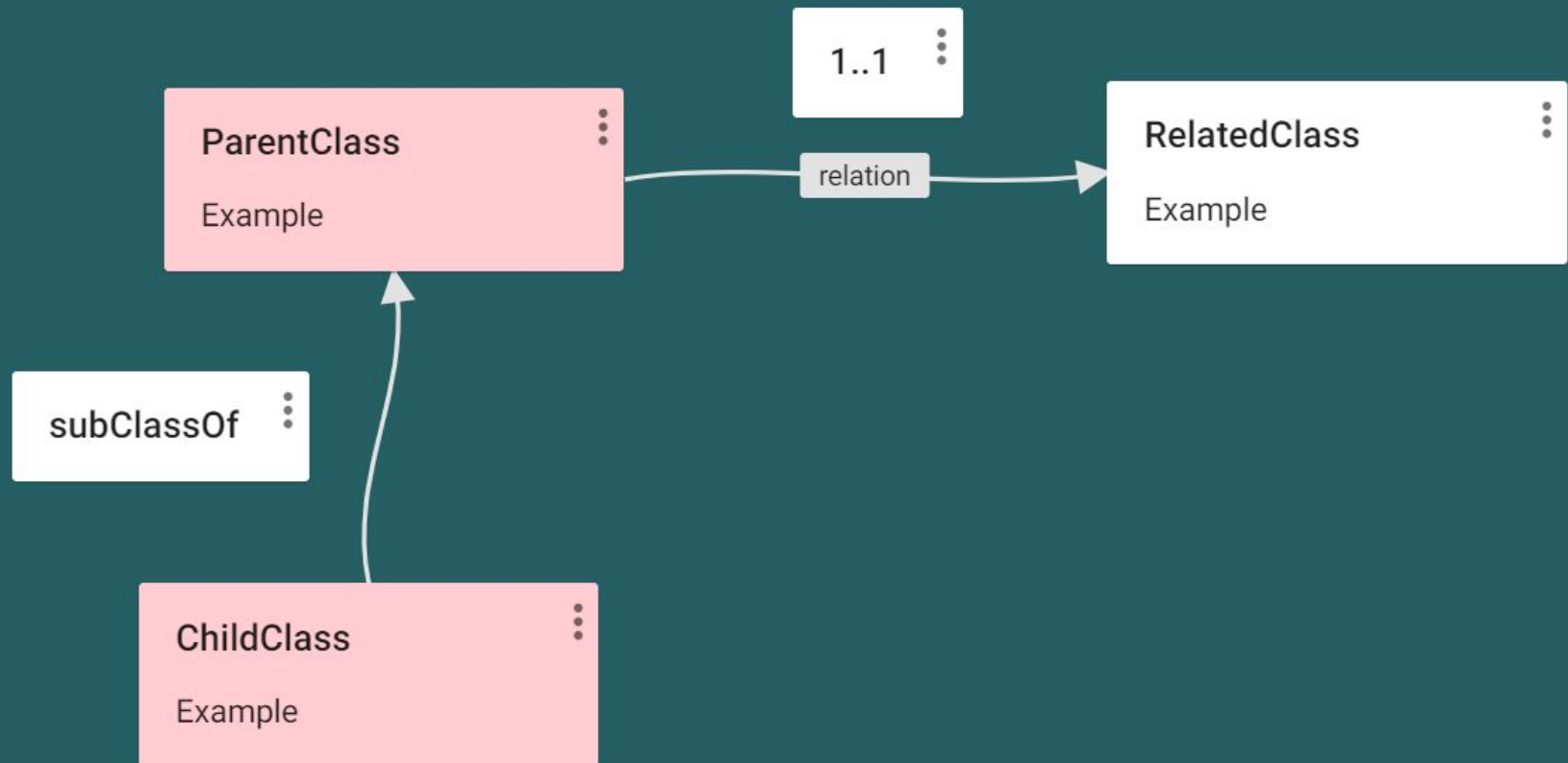


G. Use of variables by a researcher via a Variable Service

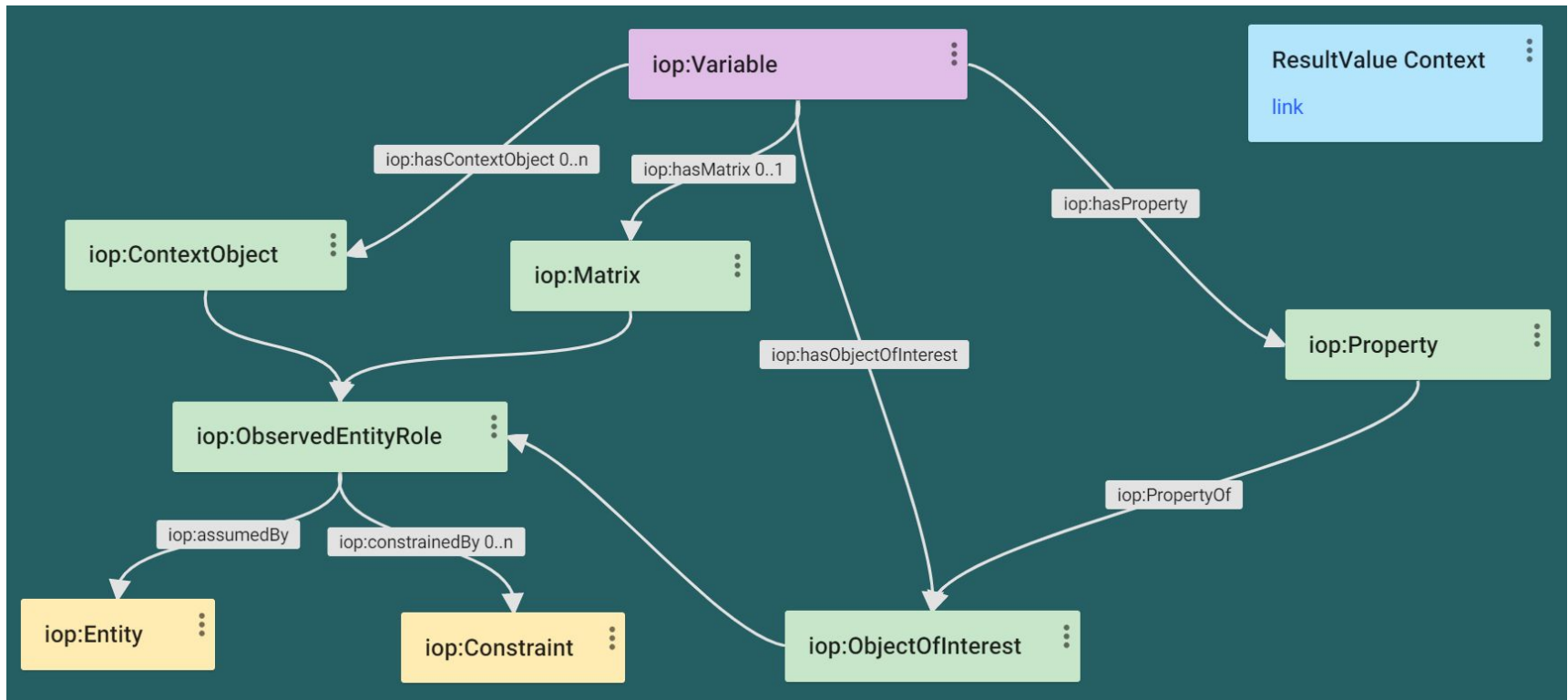


I-ADOPT Framework

How to read



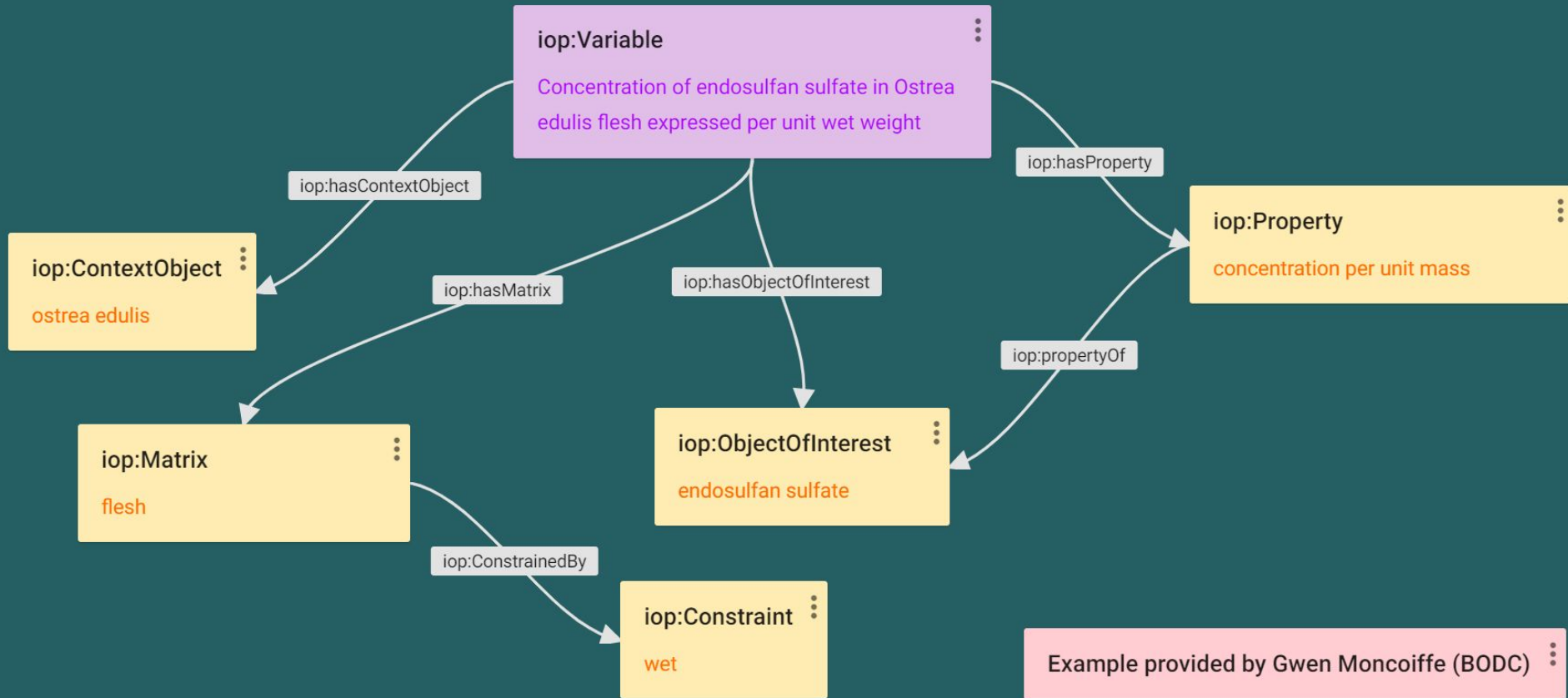
I-ADOPT Framework 1.0.0



iop:Variable: A variable is the combination of all components that are necessary to understand what was actually measured, e.g. *Concentration of endosulfan sulfate in the flesh of Ostrea edulis expressed per unit wet weight*

iop:Property: a quality of an ObjectOfinterest; thus it is a dependent element and is property of the ObjectOfInterest, e.g. *concentration, abundance, weight, presence*.

Chemistry - Concentration of substance in biota



I-ADOPT Framework – Documentation

iop:ObservedEntityRoles: Entities can assume different roles in the description of a variable.

iop:ObjectOfInterest: The entity that is the observation target, for which the property is being observed. E.g. *Endosulfan sulfate* in ‘Concentration of endosulfan sulfate in the flesh of *Ostrea edulis* expressed per unit wet weight’

iop:ContextObject: The entity that plays a descriptive role to give the context of the object of interest. E.g. *Ostrea edulis* in ‘Concentration of endosulfan sulfate in the flesh of *Ostrea edulis* expressed per unit wet weight’

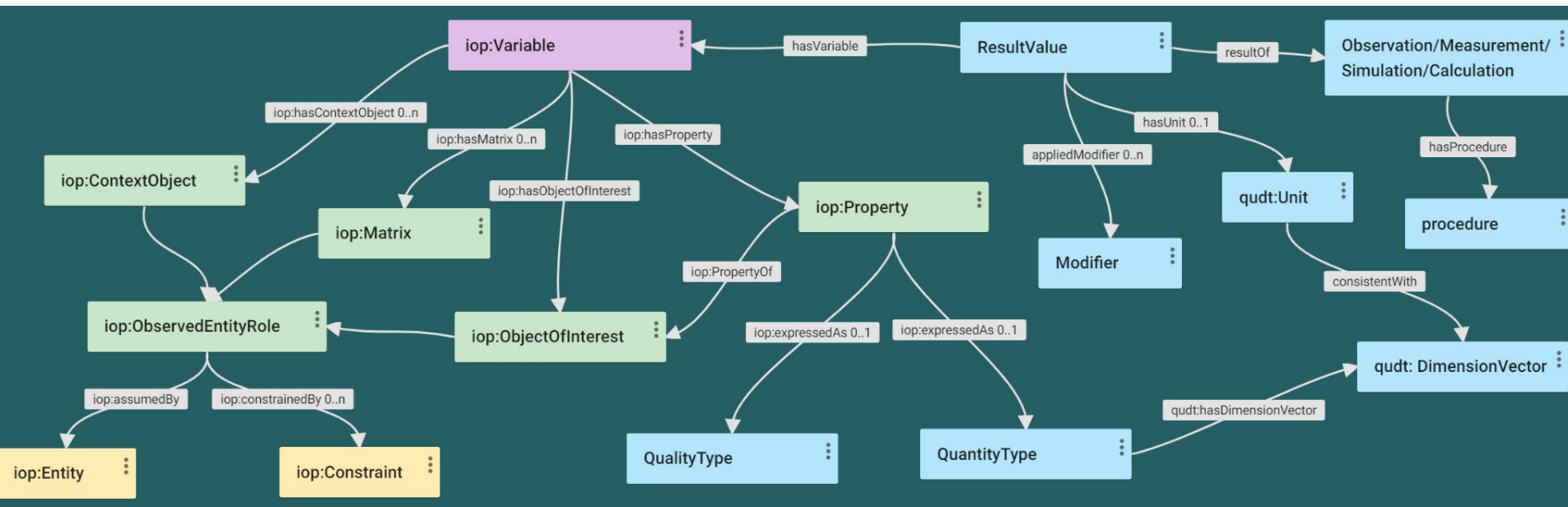
iop:Matrix: The matrix is the material context (surface or medium) of the ObjectOfInterest E.g. The *flesh* in “Concentration of endosulfan sulfate in *Ostrea edulis* flesh expressed per unit wet weight”

I-ADOPT Framework – Documentation

iop:Constraint: This may refrain any entities' conditions in their different roles as observed entities. They are in most cases measurement themselves with a set measurement result. E.g. *wet* in 'Concentration of endosulfan sulfate in the flesh of *Ostrea edulis* expressed per unit wet weight'; but also conditions like:
at 273.15 K temperature, dissolved, at 1013 hPa pressure

iop:Entities: These can be biological or non-biological entities which can play different roles in different variables' descriptions. E.g. *body, atmosphere, Nitrogen*.

I-ADOPT Framework – Context Information



ResultValue: the value that results from the event of data acquisition

Observation/Measurement/Simulation/Calculation: the data acquisition event, which could be an observation, measurement, a calculation or simulation.

Procedure: is the method of data collection

I-ADOPT Framework – Documentation

QualityType: the description type of a Property of an ObjectOfInterest represented as controlled lists of qualitative values (boolean, nominal, ordinal) for this property. Examples: *colour*, *taxon*, *presence of species*

QuantityType: the type of measurand of a Property of an ObjectOfInterest that can be measured and expressed using a number and a unit. Examples: *height*, *speed*, *concentration*, *density*. A good source for this might be qudt:quantity type

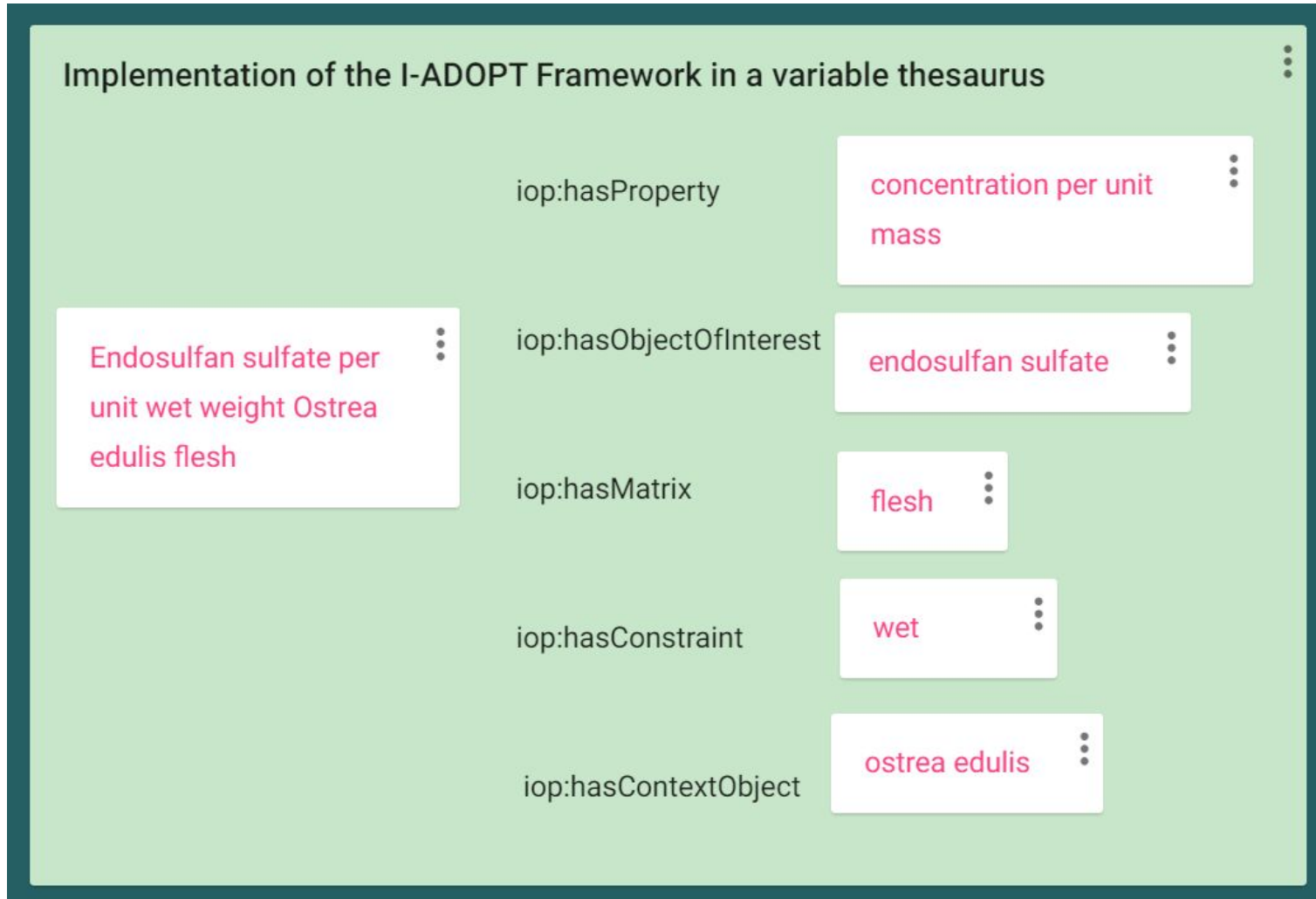
Unit: is a definite magnitude of a quantity, defined and adopted by convention or by law, that is used as a standard for measurement of the same quantity type (Wikipedia, adapted). A good source for this might be qudt:unit

DimensionVector: Each QuantityType has a DimensionVector that describe it in terms of 7 basic dimensions, and the associated base units (QUDT.org). A good source for this might be qudt:dimension vector

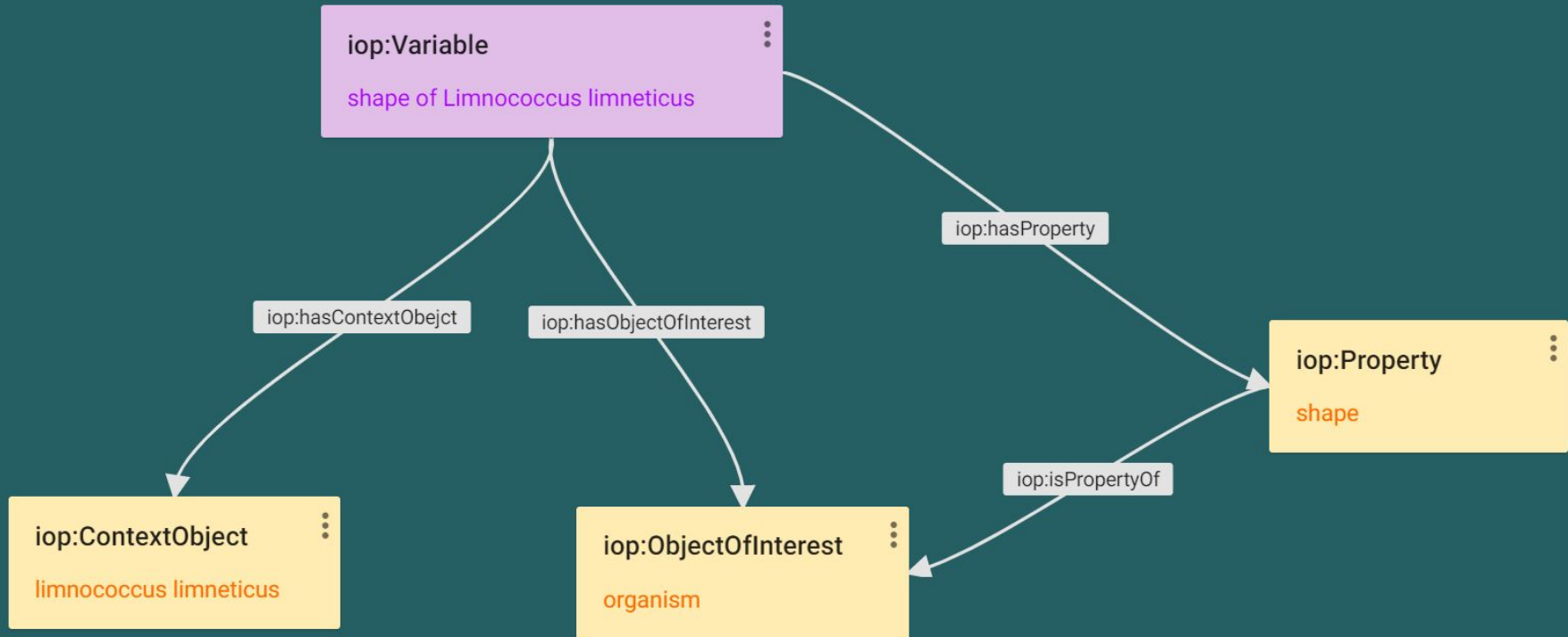
Modifier: Mathematical or qualitative transformation or operation done on the result of a measurement result of a quantifiable property. E.g. *standard deviation*, *average*.

Chemistry - Concentration of substance in biota

SKOS implementation



Biodiversity example: Shape of organism



Biodiversity example provided by Ilaria Rosati (LifeWatch Italy)

Biodiversity example: Shape of organism

SKOS implementation

Implementation of the I-ADOPT Framework in a biodiversity thesaurus (e.g. LifeWatch vocabulary)

LifeWatch thesaurus

shape of *Limnococcus*
limneticus in water

iop:hasProperty

shape

<http://thesauri.lifewatchitaly.eu/PhytoTraits/index.php?tema=23&/shape>

iop:hasObjectOfInterest

organism

http://www.ontobee.org/ontology/OBI?iri=http://purl.obolibrary.org/obo/OBI_0100026

iop:hasContextObject

Limnococcus limneticus

<http://www.marinespecies.org/aphia.php?p=taxdetails&id=576250>

Addressing the semantic alignment and data integration use cases: example from the marine domain

Explaining the Interoperability Framework using
examples of Essential Climate, Ocean and
Biodiversity Variables

Gwenaëlle Moncoiffé



☐  6 Open ✓ 0 Closed Author ▾ Label ▾ Projects ▾

☐  **Researcher- find long time series data - efficient semantic search across multiple vocabularies**

facet semantic data search semantic alignment

#20 opened on Apr 7, 2020 by mabablue

☐  **Data manager - interchange between BODC PUV P01 and CF standard names - deliver SeaDataNet CF-NetCDF** data annotation data integration semantic alignment

#17 opened on Feb 26, 2020 by louatbodc

☐  **Terminology provider - access agreed mappings - gain efficiency in data exchange**

data model alignment semantic alignment terminology management

#15 opened on Jan 11, 2020 by gwemon

☐  **Bio-loggingData - NormalizeDataMeasurements - Interoperability** data integration

facet semantic data search semantic alignment

#10 opened on Jan 6, 2020 by sarahcd

☐  **Vocabulary manager - Select the right URI for "nitrogen" - Better data integration and interoperability** semantic alignment terminology search

#6 opened on Nov 19, 2019 by jonquet

☐  **Soil Data manager - Normalize data - Achievement of a standard** data model alignment

semantic alignment semantic modelling

Parameter labelling schemes – 3 examples



CF MetaData

CF
Standard
Names

See: <https://cfconventions.org/Data/cf-standard-names/docs/guidelines.html>

P07 - Climate and Forecast Standard Names

mole_concentration_of_dissolved_molecular_oxygen_in_sea_water

CF14N29

Mole concentration means number of moles per unit volume, also called "molarity", and is used in the construction mole_concentration_of_X_in_Y, where X is a material constituent of Y. A chemical or biological species denoted by X may be described by a single term such as "nitrogen" or a phrase such as "nox_expressed_as_nitrogen". The chemical formula for molecular oxygen is O2.



BODC P01
Vocabulary
collection
(PUV)

P01 - BODC Parameter Usage Vocabulary

Concentration of oxygen {O2 CAS 7782-44-7} per unit volume of the water body [dissolved plus reactive particulate phase]

DOXYZZXX

WC_dissO2

Concentration of dissolved oxygen per unit volume of the water column. Oxygen may be expressed in terms of mass, volume or quantity of substance.

See: <https://github.com/nvs-vocab/P01>

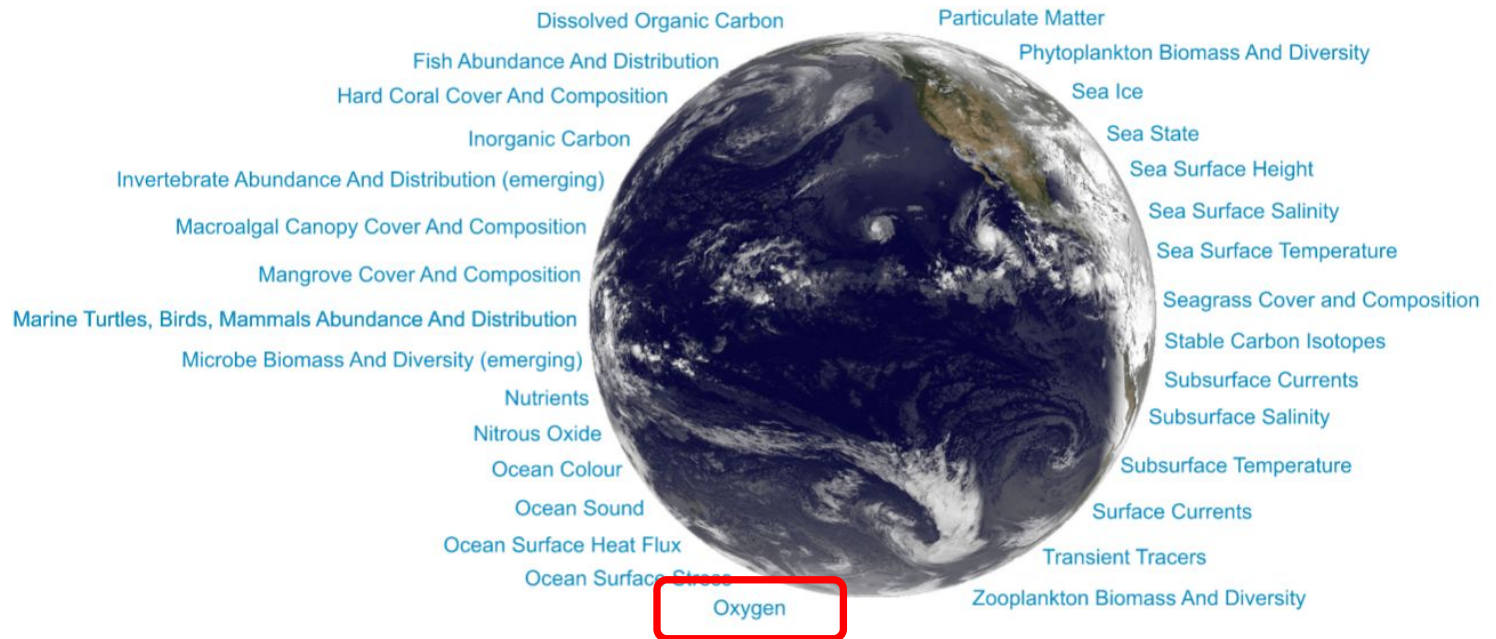


ICES
PARAM
collection

Code:	DOXY
Description:	dissolved oxygen
Long Description:	
CodeType:	PARAM - Parameters
Deprecated:	False
Created Date:	2006-10-23
Modified Date:	2006-10-23
Link to this Code:	//vocab.ices.dk/?CodeID=33506

See: <https://www.ices.dk/data/vocabularies/Pages/default.aspx>

We need access to data – example GOOS EOVs



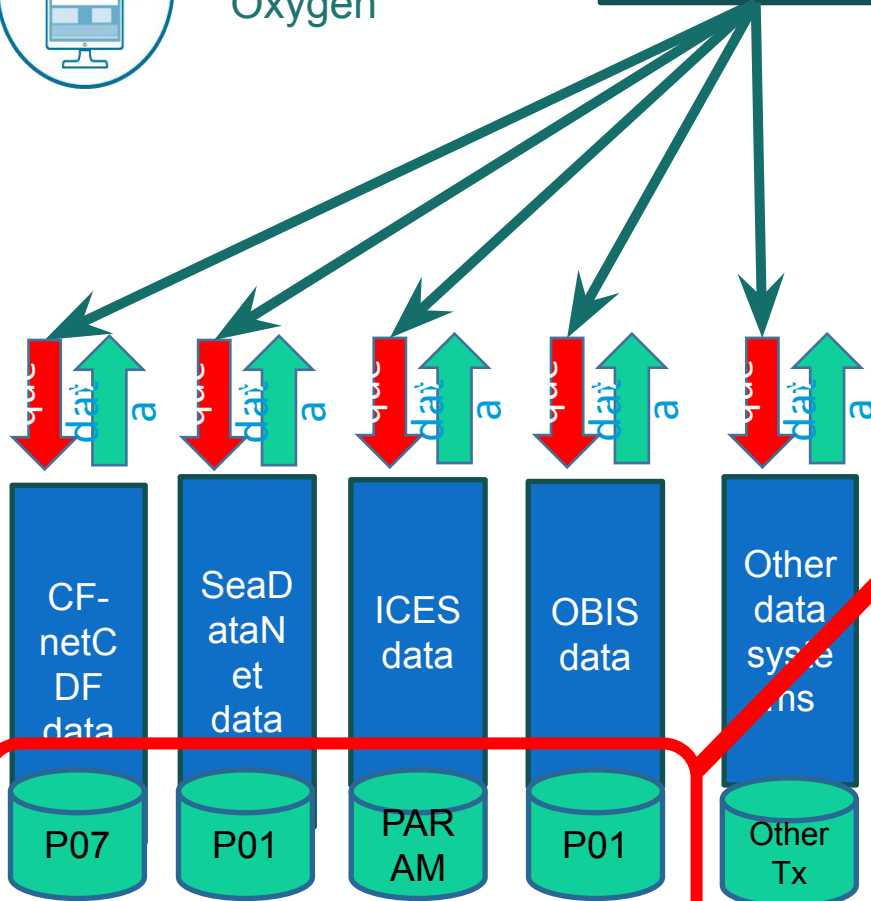
Oxygen?

Searching for dissolved oxygen data across repositories



User requests
EOV
"Oxygen"

Select EOV
parameter



Structured?



Machine
readable?



Interoperable?



Dissolved oxygen in CF standard name, BODC PUV, and ICES PARAM - a selection

CF Standard Names

P07 - Climate and Forecast Standard Names

mole_concentration_of_dissolved_molecular_oxygen_in_sea_water

CF14N29

Mole concentration means number of moles per unit volume, also called "molarity", and is used in the construction mole_concentration_of_X_in_Y, where X is a material constituent of Y. A chemical or biological species denoted by X may be described by a single term such as "nitrogen" or a phrase such as "nox_expressed_as_nitrogen". The chemical formula for molecular oxygen is O2.

BODC P01 Vocabulary collection (PUV)

P01 - BODC Parameter Usage Vocabulary

Concentration of oxygen (O2 CAS 7782-44-7) per unit volume of the water body [dissolved plus reactive particulate phase]

DOXYZZXX

WC_dissO2

Concentration of dissolved oxygen per unit volume of the water column. Oxygen may be expressed in terms of mass, volume or quantity of substance.

ICES PARAM collection

Code: DOXY
Description: dissolved oxygen
Long Description:

CodeType: [PARAM](#) - Parameters

Deprecated: False

Created Date: 2006-10-23

Modified Date: 2006-10-23

Link to this Code: [//vocab.ices.dk/?CodeID=33506](http://vocab.ices.dk/?CodeID=33506)

Are they the same?



P07 - Climate and Forecast Standard Names

mole_concentration_of_dissolved_molecular_oxygen_in_sea_water

CF14N29

Mole concentration means number of moles per unit volume, also called "molarity", and is used in the construction mole_concentration_of_X_in_Y, where X is a material constituent of Y. A chemical or biological species denoted by X may be described by a single term such as "nitrogen" or a phrase such as "nox_expressed_as_nitrogen". The chemical formula for molecular oxygen is O2.

P01 - BODC Parameter Usage Vocabulary

Concentration of oxygen {O2 CAS 7782-44-7} per unit volume of the water body [dissolved plus reactive particulate phase]

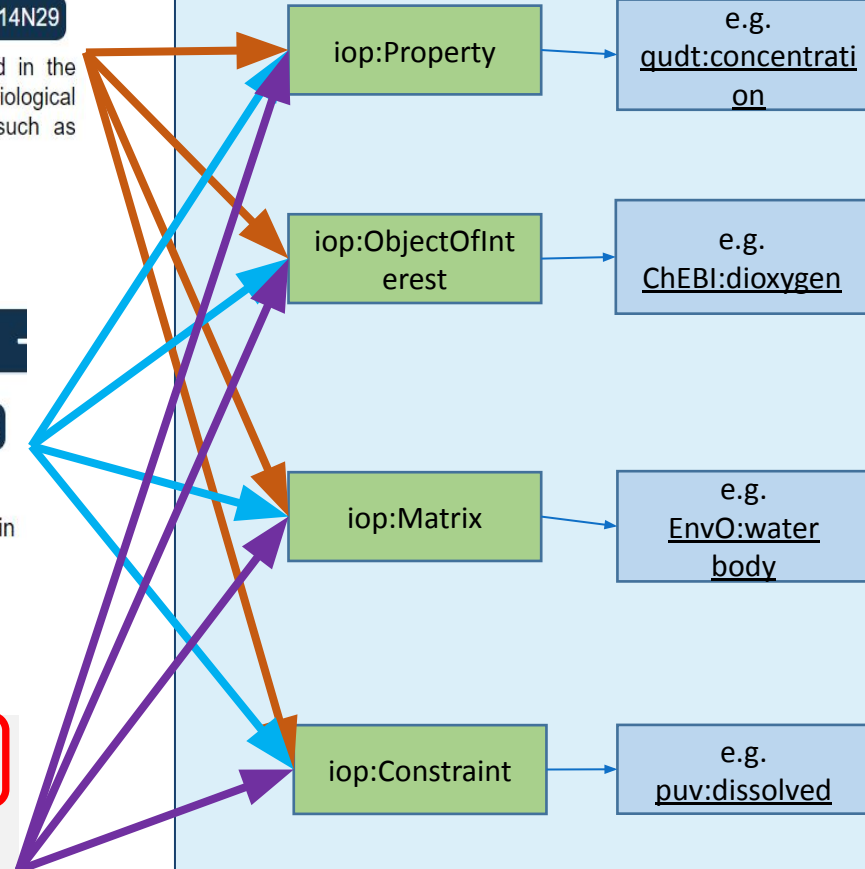
DOXYZZXX

WC_dissO2

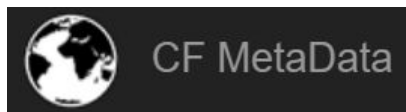
Concentration of dissolved oxygen per unit volume of the water column. Oxygen may be expressed in terms of mass, volume or quantity of substance.



Code:	DOXY
Description:	dissolved oxygen
Long Description:	
CodeType:	PARAM - Parameters
Deprecated:	False
Created Date:	2006-10-23
Modified Date:	2006-10-23
Link to this Code:	//vocab.ices.dk/?CodeID=33506



I-ADOPT Interoperability Framework
with deferenceable components using a common set of interlinked "FAIR" terminologies and ontologies



Parameter Labelling Schemes

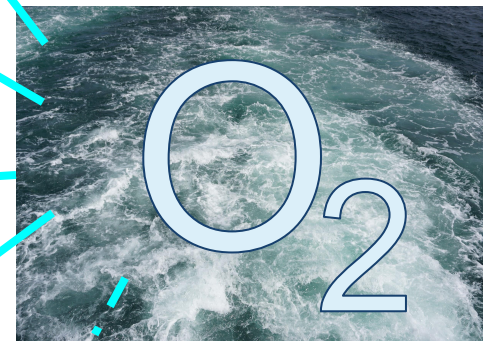
iop:Property is
"Concentration"

iop:objectOfInter
est is "molecular
oxygen"

iop:matrix is
"water body"

iop:constraint is
"dissolved"

I-ADOPT Interoperability
Framework

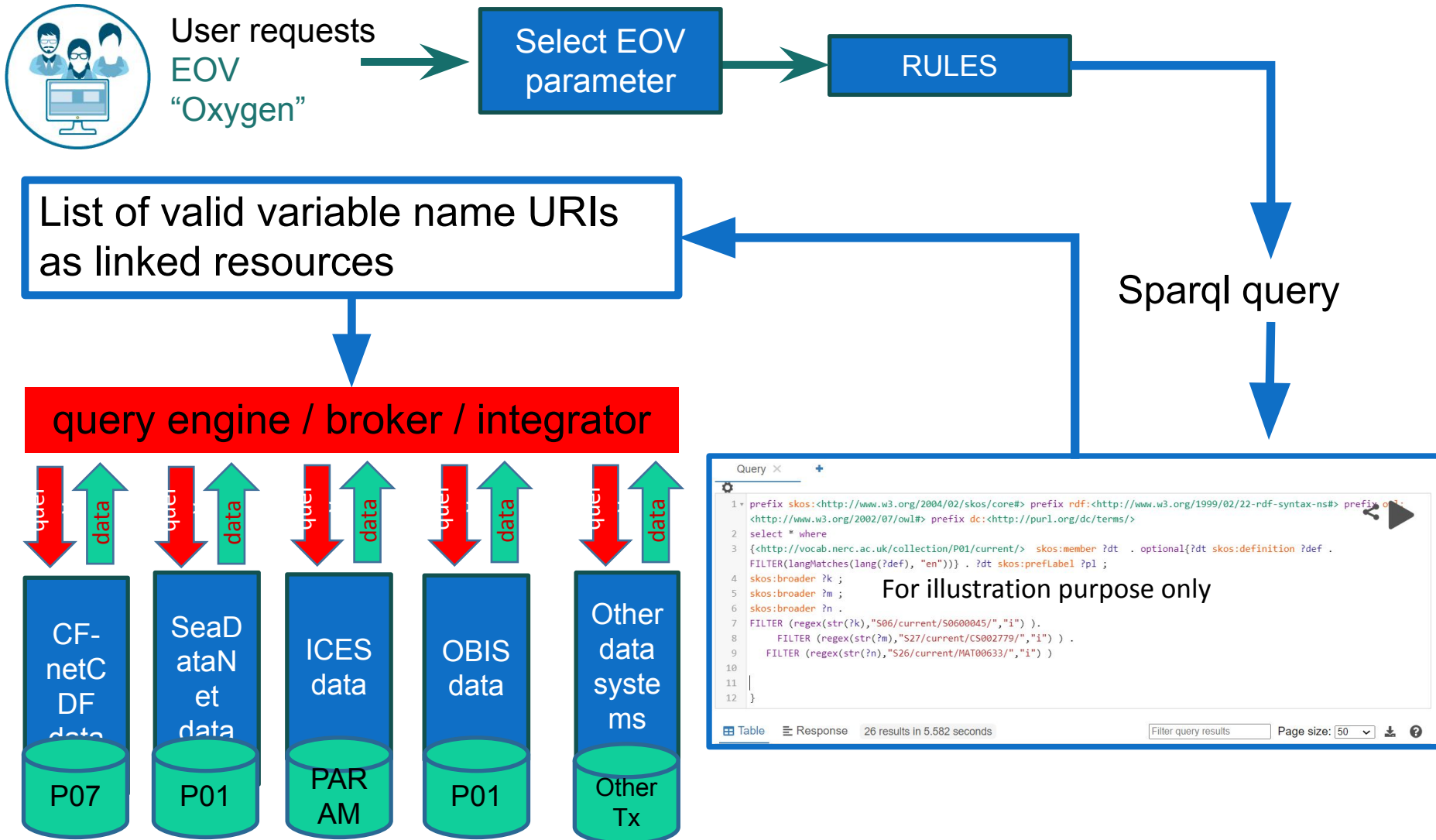


EOV Concept
"Oxygen"

Units

Units and
Dimensions
schemes like e.g.
qudt

Example of implementation



I-ADOPT Framework – Examples

Chemistry example: [Endosulfan example](#)
with a possible SKOS [implementation](#)

Biodiversity example: [Shape of organism example](#)

Earth observation example: [Taxonomy of grass](#)

Life science example: [Molecular example](#)

Social sciences example: [Survey question](#)

Thank you for your interest! Join [I-ADOPT](#) if interested!

Check out our resources:

I-ADOPT WG [wiki on RDA](#)

I-ADOPT Framework [documentation](#)

I-ADOPT Framework [padlet](#)

I-ADOPT on [GitHub](#)

Any questions?