Representation in RDF of Scientific Variables

Contact : Pascal.Neveu@inrae.fr

Variable

Une variable est une caractéristique commune à un ensemble d'entités impliquées dans une expérimentation. La valeur de cette caractéristique varie entre les entités. Cette valeur est issue d'une observation ou d'une mesure ou d'une simulation ou d'un calcul effectué sur une entité identifiée individuellement (site, objet scientifique, parcelle…) sous certaines conditions (méthode de mesure, niveau d'agrégation…).

Par exemple, la variable "circonférence des arbres" s'effectue sur le tronc (une partie de l'objet scientifique) d'un arbre identifié à l'aide d'un compas forestier à une hauteur de 1,30 m. Autre exemple, la variable "vitesse du vent" s'effectue sur l'entité vent par une station météo sur un site identifié et à l'aide d'un anémomètre.

Discrete and Continuous Data

Discrete data can only take on certain individual values.

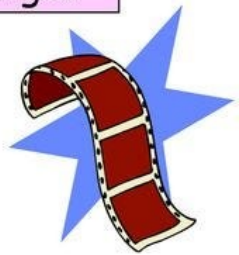
Continuous data can take on any value in a certain range.

Example 1

Number of pages in a book is a **discrete variable.**

Example 2

Length of a film is a **continuous variable.**



Example 3

Shoe size is a **Discrete variable.** E.g. 5, $5\frac{1}{2}$, 6, $6\frac{1}{2}$ etc. Not in between.

Example 4

Temperature is a **continuous variable.**

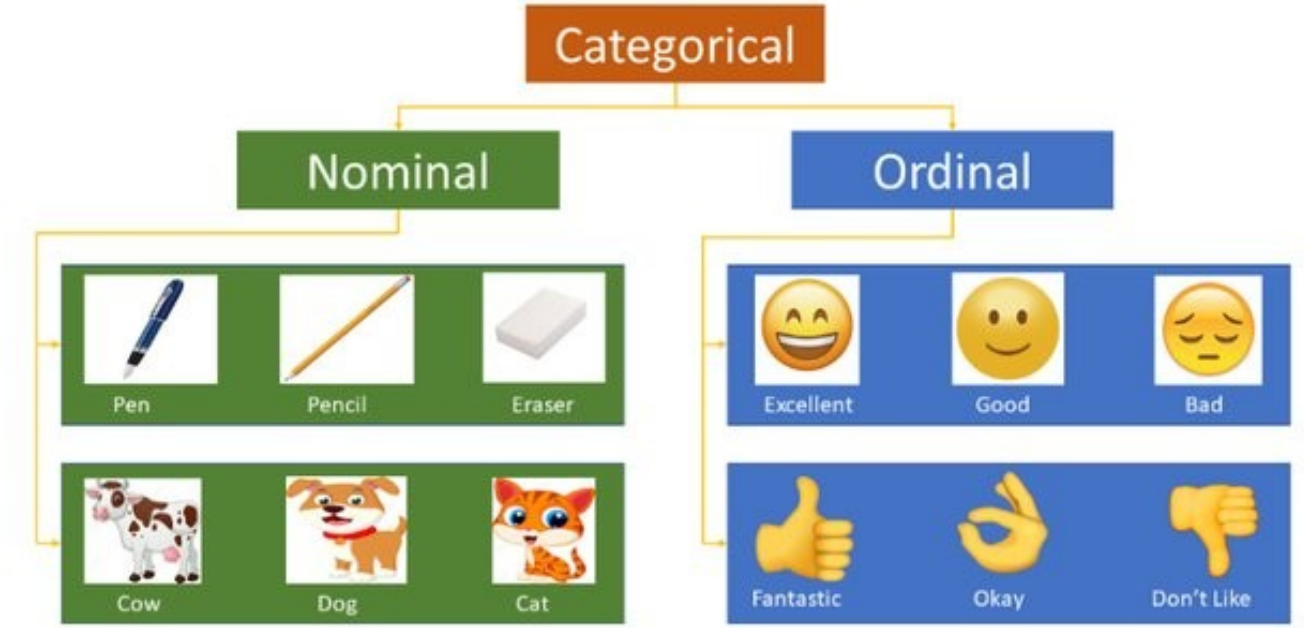
Example 5

Number of people in a race is a **discrete variable.**

Example 6

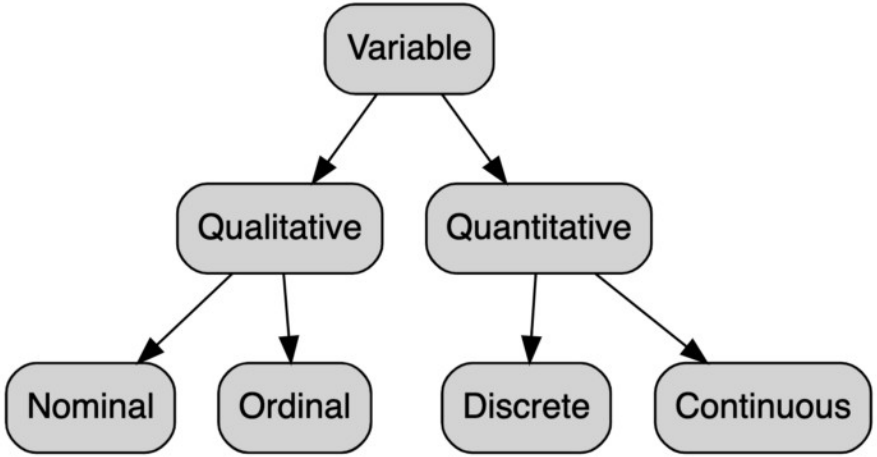
Time taken to run a race is a **continuous variable**

Qualitative Variable

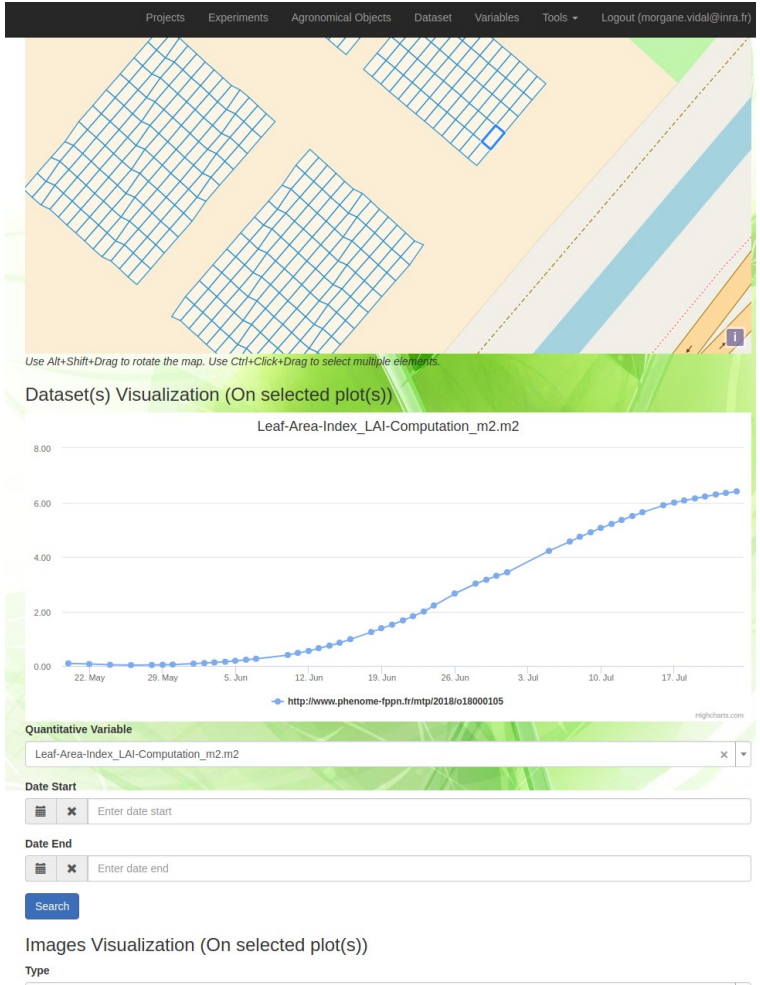


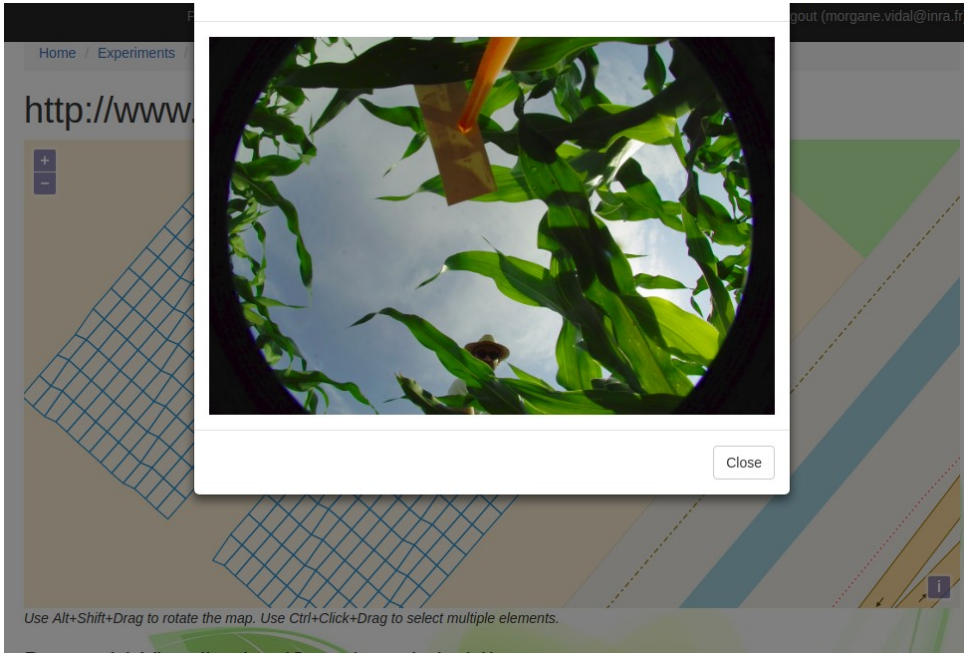
Variable

Variable is the key element of data structuring and analytics



Phenomics Variable Example





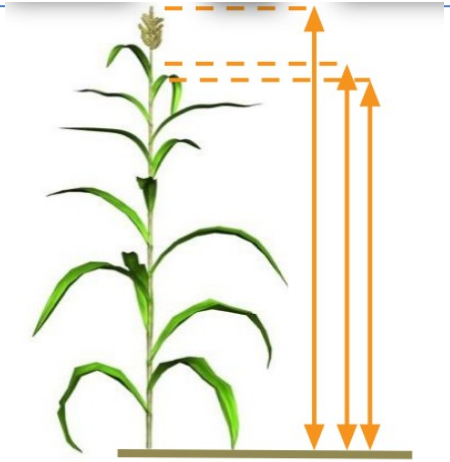
What Can I do?

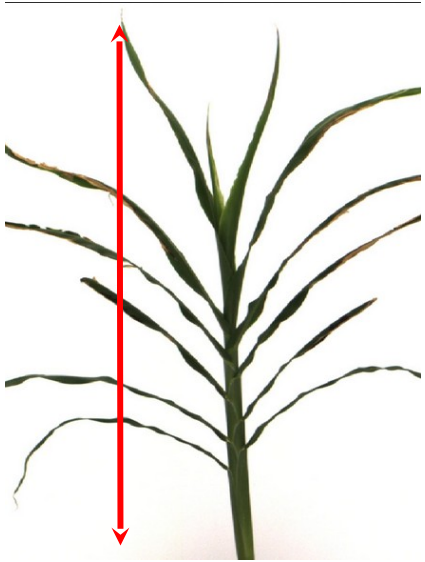
	CodeDpt	sexe	jour	Effect1	Effec2	r
2	1	0	2020-03-18	16	0	
3	1	1	2020-03-18	14	0	
4	2	1	2020-03-18	10	0	
5	3	0	2020-03-18	34	0	
6	3	1	2020-03-18	38	0	
7	4	0	2020-03-18	0	33	
8	4	1	2020-03-18	0	27	

	A	B	C	D	E	F	G
1	dep	sexe	jour	hosp	rea	rad	dc
2	1	0	2020-03-18	2	0	1	0
3	1	1	2020-03-18	1	0	1	0
4	1	2	2020-03-18	1	0	0	0
5	2	0	2020-03-18	41	10	18	11
6	2	1	2020-03-18	19	4	11	6
7	2	2	2020-03-18	22	6	7	5
8	3	0	2020-03-18	4	0	1	0
9	3	1	2020-03-18	1	0	0	0
10	3	2	2020-03-18	3	0	1	0
11	4	0	2020-03-18	3	1	2	0
12	4	1	2020-03-18	3	1	0	0
13	4	2	2020-03-18	0	0	2	0
14	5	0	2020-03-18	8	1	9	0

The reality is too often

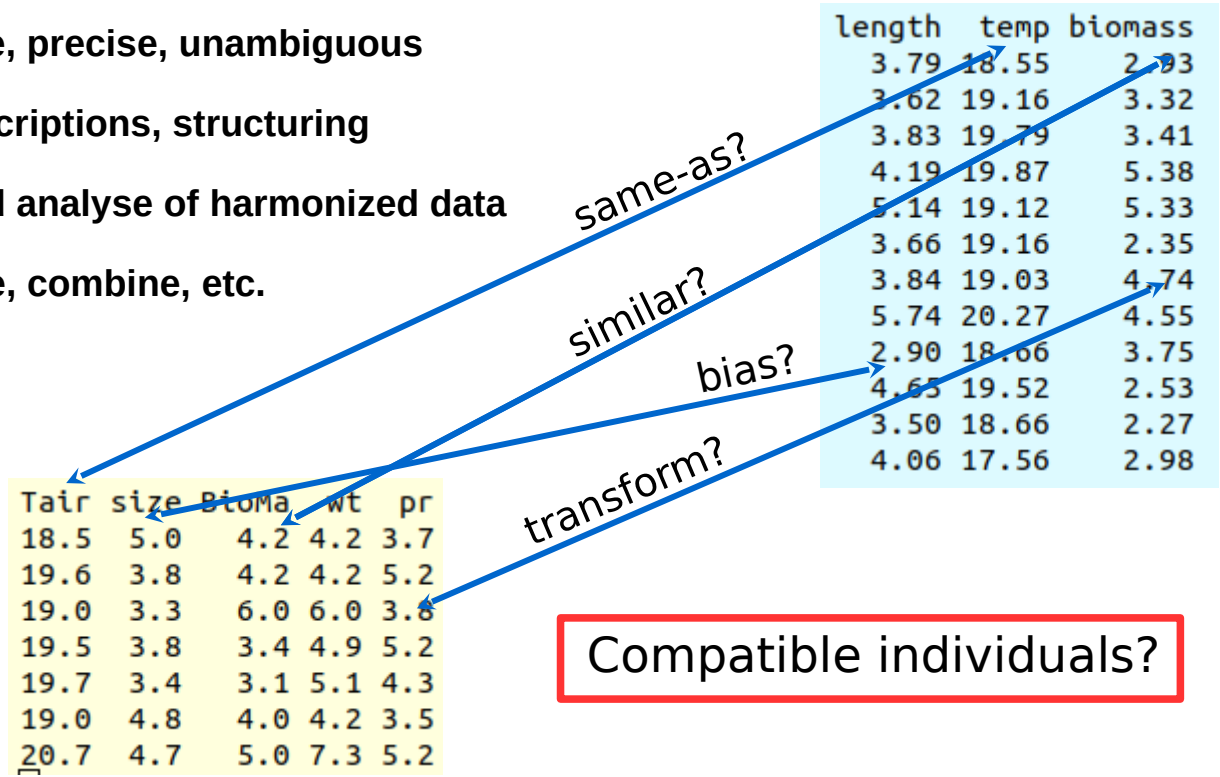
- **No variable ID**
- Same name for different variables
- Different names for same variable
- No (or not machine readable) variable description
- Unstable variable





Formalized variable with ID

- **Aims FAIR data with a focus on interoperability**
- **Share an approach: simple, precise, unambiguous**
- **Share/Reuse variable: descriptions, structuring**
- **Make easier aggregate and analyse of harmonized data**
- **Be able to compare, merge, combine, etc.**



Variable Model

Requires to formalize variables

- **that anyone could easily understand**
- **that can be manipulate with machine**

A Common representation in biology

Trait characteristic of an organism

Method is the way we have to measure variable. Giving description of protocols and type of instruments used.

Scale/Unit allow to specify unit or scale use for ordinal variable

Variable Model

We need to create an infinite number of variables!

Entity is the entity of interest of the variable, what we are measuring on.

Characteristic is the observed property such as physical quantity or biological quality.

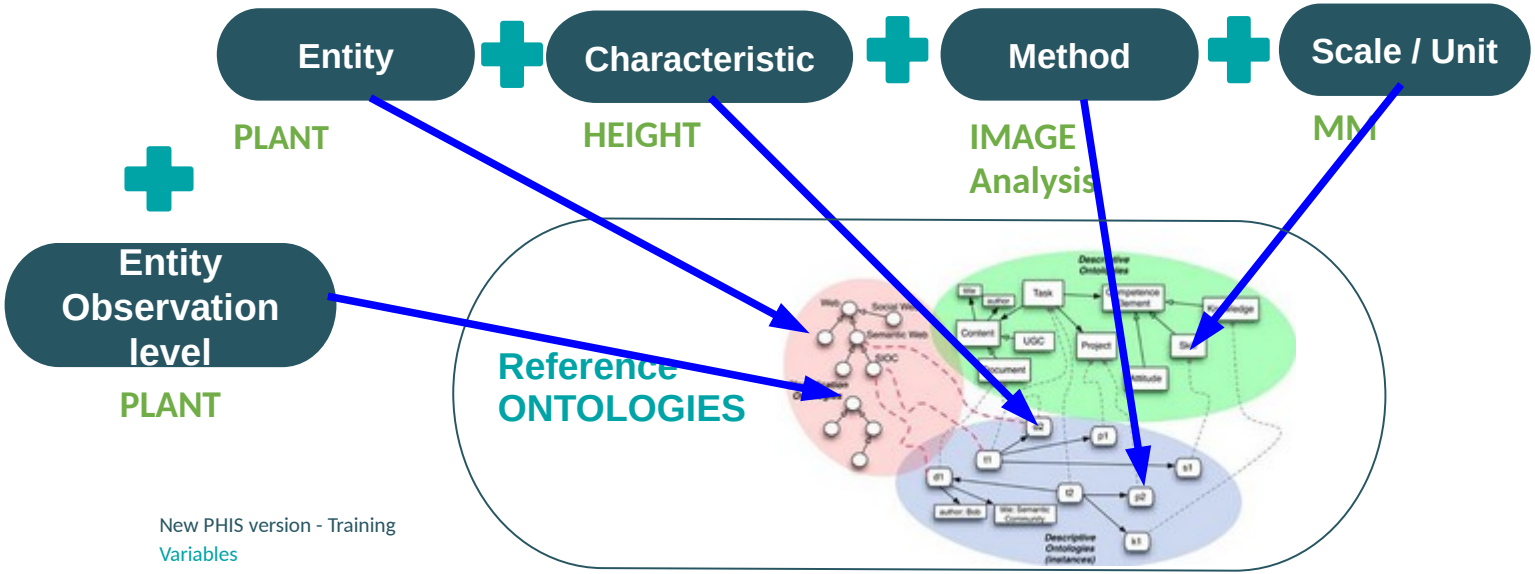
Observation level entity : entity of the level of individual measurements or aggregations

Method is the way we have to measure or aggregate variable. Giving description of protocols and type of instruments used.

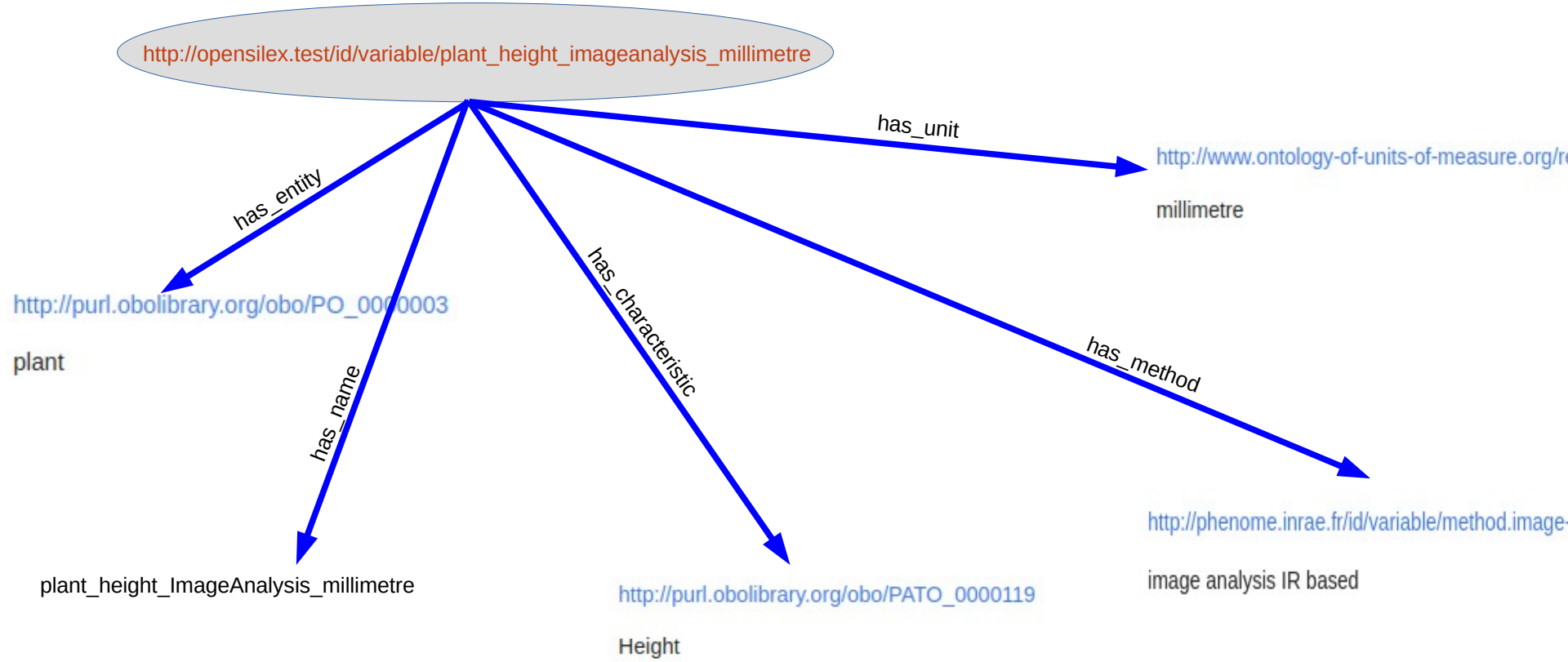
Scale/Unit allow to specify unit or scale use for ordinal variable

Variable Model

- Enabling semantically precise descriptions
- Decomposing description into standardized elements
- Link to existing vocabularies/ontologies
- Make description machine readable

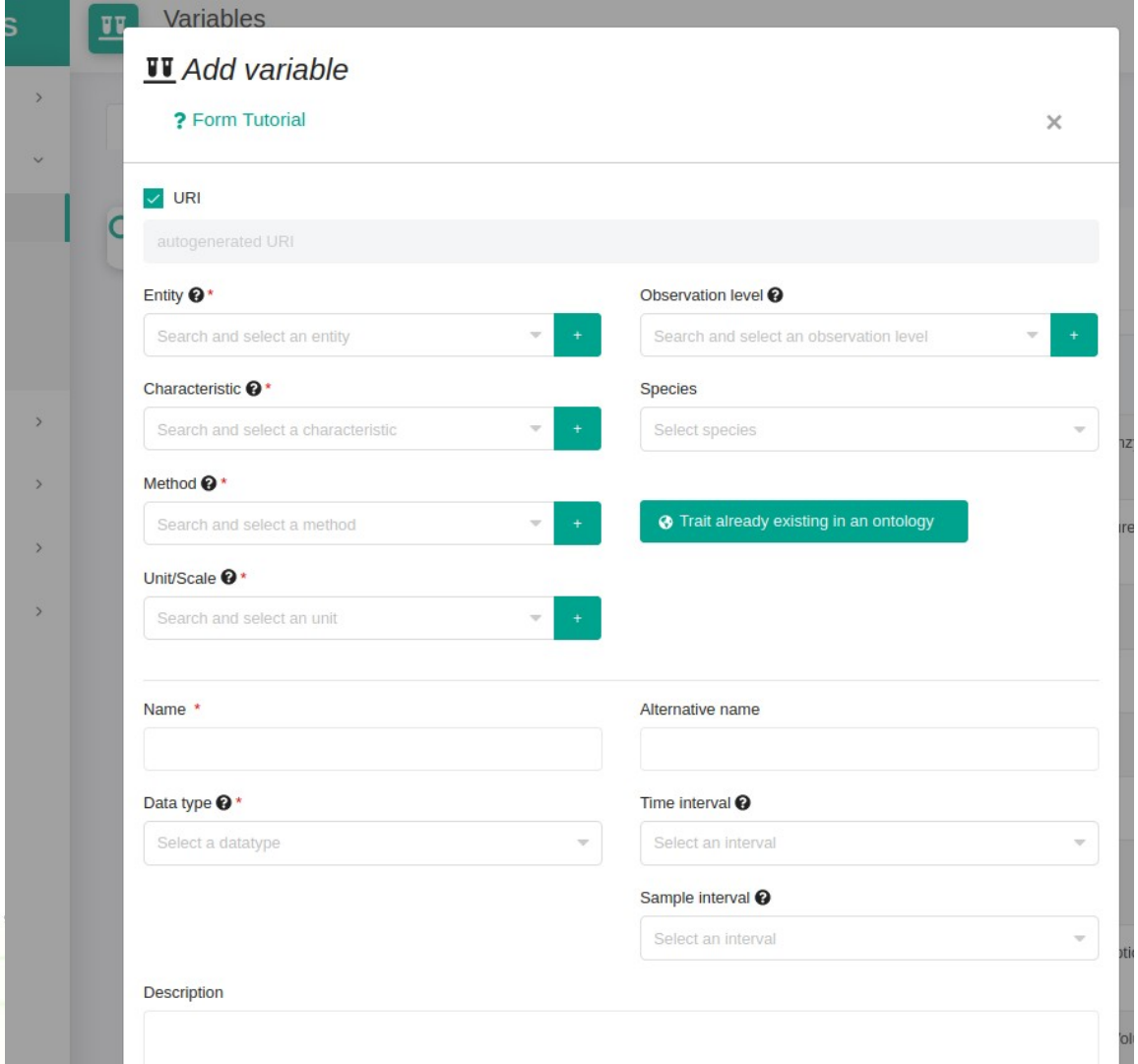


Variable Class

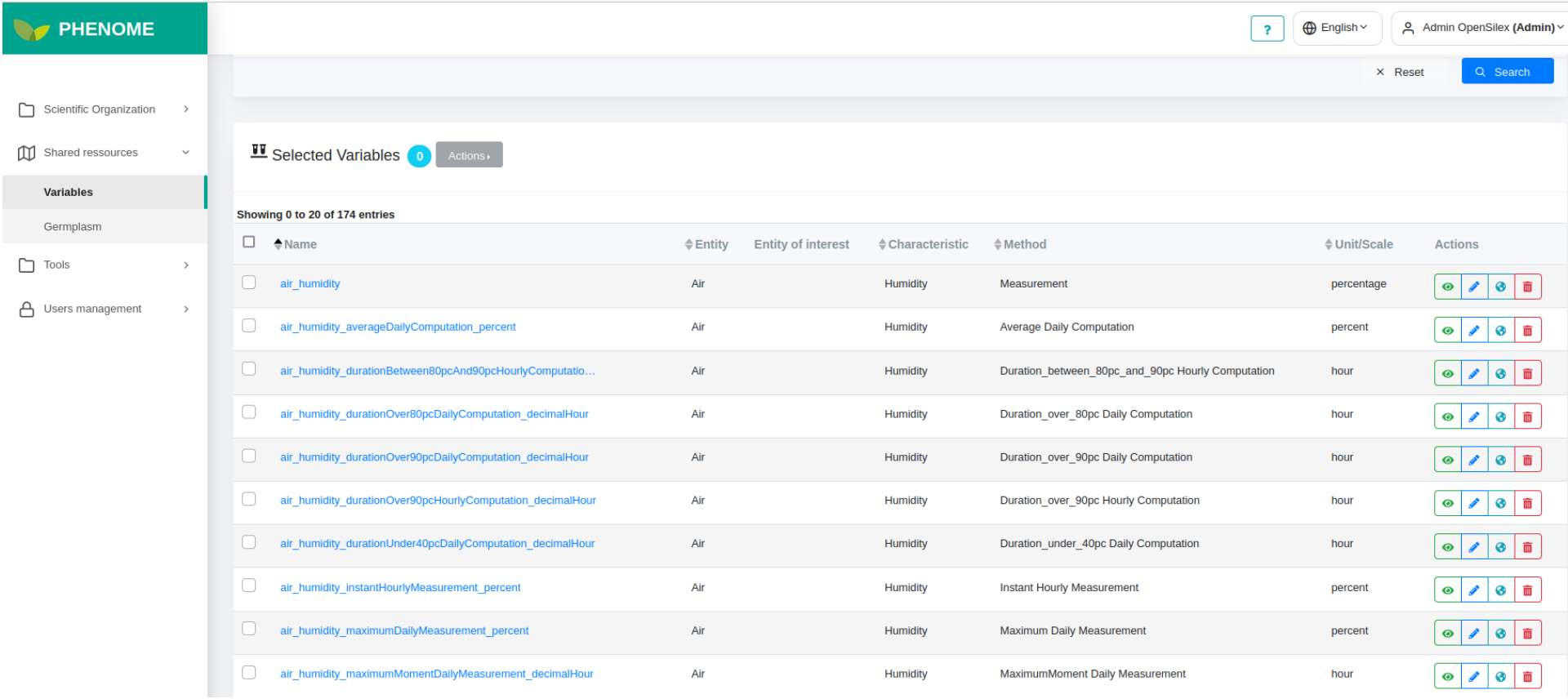


Needs of flexibility

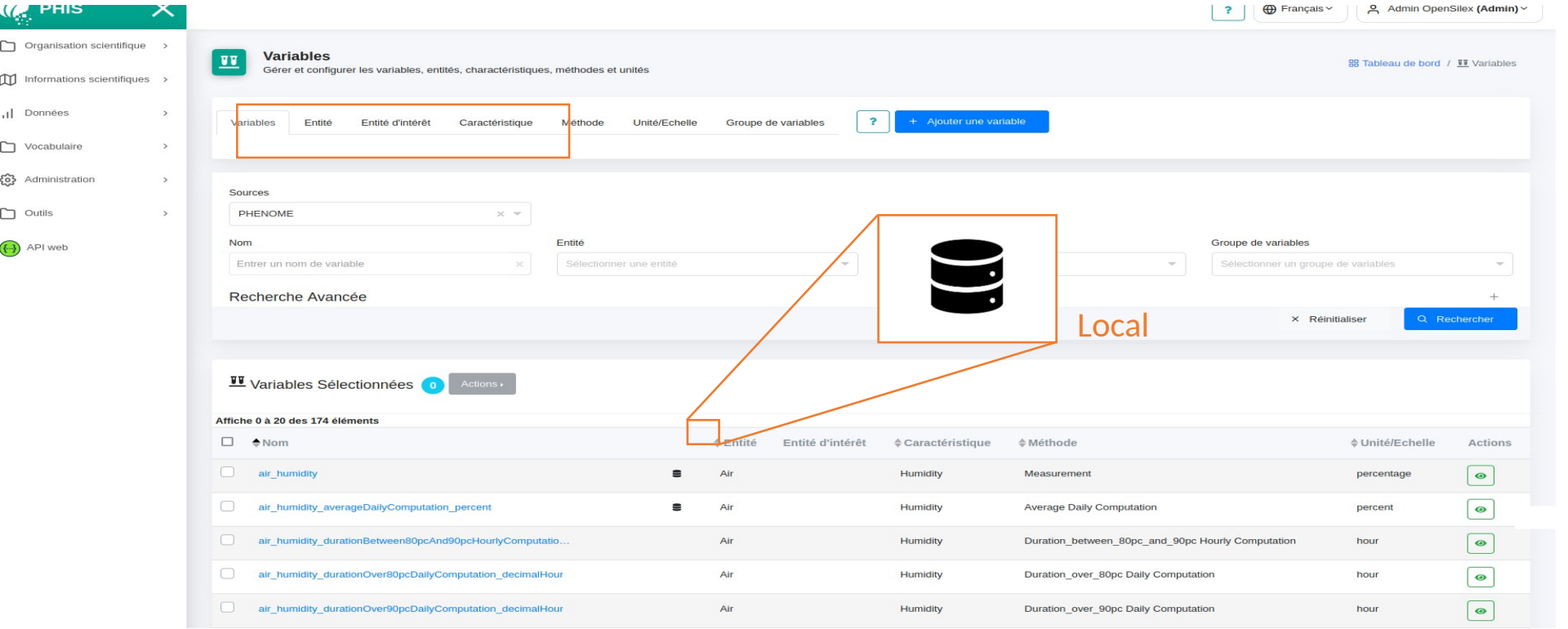
- Variable declaration
- Interoperability



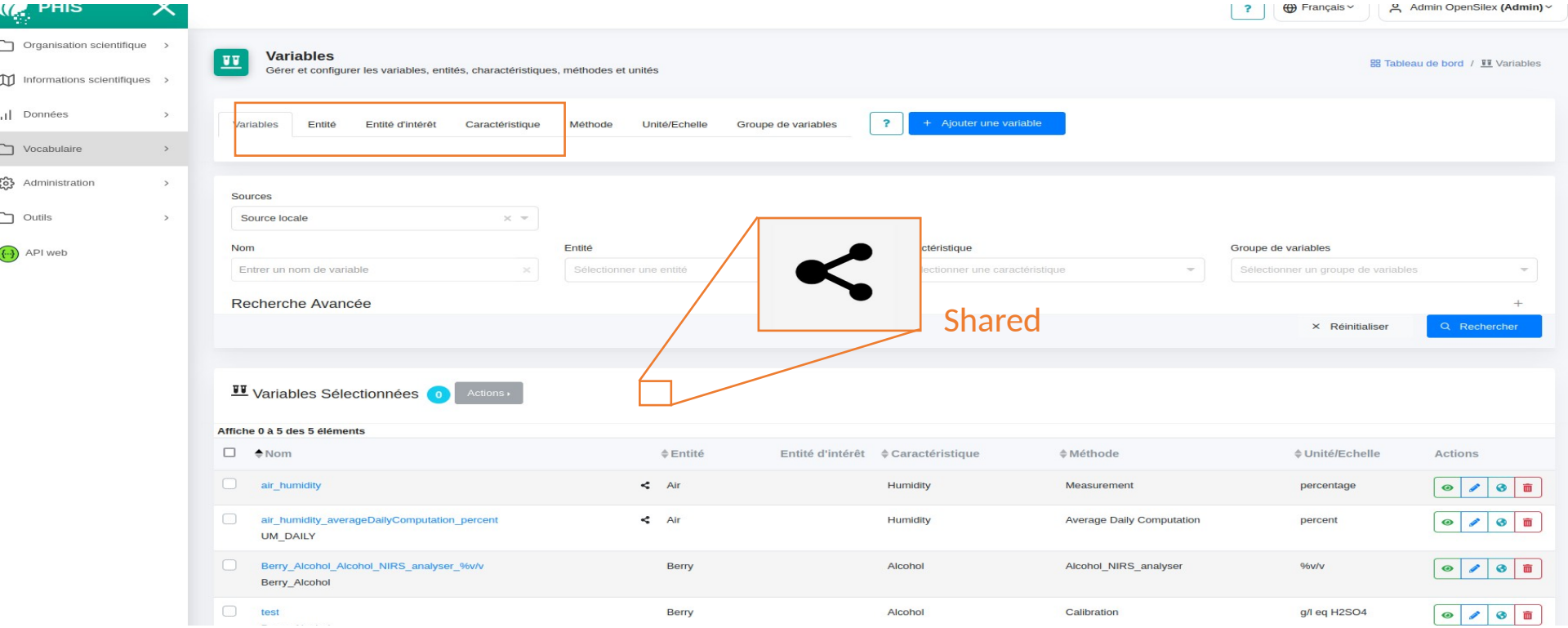
Share and Reuse variables



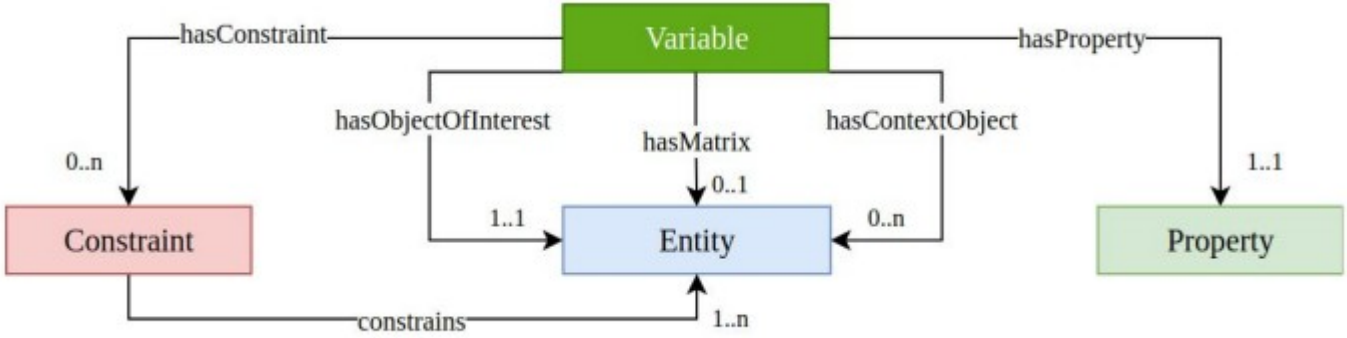
Share and Reuse variables



Share and Reuse variables



I-ADOPT Model



- **Object of interest = Entity**
- **Property = Characteristic**
- Matrix = **Level of observation entity?**
- **Context object = Level of observation entity?**
- Constraint =

Example I-ADOPT

The measurement of the molar conductivity of potassium iodide in a solution of glycerol and water (at different temperatures and concentrations of glycerol)

Variable: molar conductivity (property) **OF** potassium iodide (ObjectOfInterest) **IN** glycerol water mixture

Matrix: glycerol + water (different compositions) + Property: molar conductivity
Constraints: pressure of the system, system temperature, glycerol concentration, potassium iodide conc. (by calculation the concentration of water)

Entities: mixtures of the potassium iodide, glycerol and water at specific concentrations

ContextObject: glycerol/water mixture which can be tested to get the 'background' molar conductivity

constrains: Entity (mixture at a specific set of concentrations) constrains Variable (molar conductivity of KI in glycerol/water)

hasConstraint: Variable (molar conductivity of KI in glycerol/water) hasConstraint Entity (mixture at a specific set of concentrations)

hasContextObject: Variable (molar conductivity of KI in glycerol/water) hasContextObject ContextObject (background molar conductivity measurement of a mixture of glycerol/water with specific composition)

hasMatrix: Variable (molar conductivity of KI in glycerol/water) hasMatrix Matrix (mixture of glycerol/water with specific composition that the potassium iodide is dissolved in)

hasObjectOfInterest: Variable (molar conductivity of KI in glycerol/water) hasObjectOfInterest Entity (potassium iodide)

hasProperty: Variable (molar conductivity of KI in glycerol/water) has Property Property (molar conductivity)