

# Thésaurus Theia/OZCAR: Implémentation de l'ontologie I-ADOPT pour faciliter l'interopérabilité des variables scientifique

Charly Coussot (1), Isabelle Braud (2), Véronique Chaffard (3), Brice Boudevillain (4), Sylvie Galle (3)

1. Université Grenoble Alpes, IRD, CNRS, Météo-France, INRAE, OSUG, 38000 Grenoble, France

2. INRAE, RiverLy, Villeurbanne, France

3. Université Grenoble Alpes, IRD, CNRS, Grenoble-INP, IGE, 38000 Grenoble, France

4. Université Grenoble Alpes, CNRS, IRD, Grenoble-INP, IGE, 38000 Grenoble, France

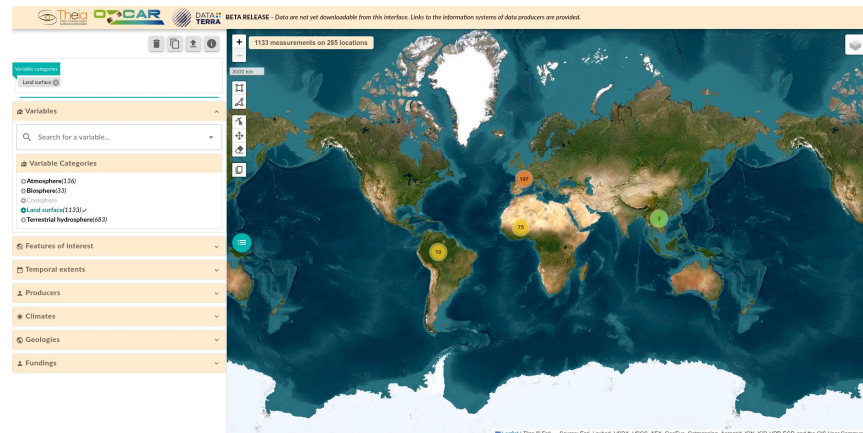
## Partage des données dans le cadre de Theia/OZCAR et de la science interdisciplinaire



Sources de données : 22 observatoires de long terme français de zones critiques, avec 22 systèmes d'information préexistants.

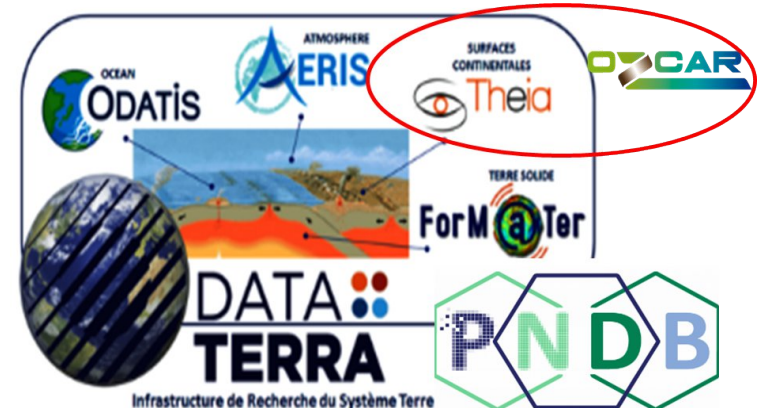
Grande hétérogénéité des données (diversité des noms de variables), de la gestion des données et des pratiques.

Les données doivent être réutilisées dans un contexte de recherche interdisciplinaire



Système d'information Theia/OZCAR

 **OneWater**  
Eau Bien Commun



# Un vocabulaire contrôlé pour soutenir les objectifs de le SI Theia/OZCAR

## Objectifs du SI Theia/OZCAR:

- Exposer des données provenant de sources hétérogènes sur [un portail de données unique](#) avec des métadonnées homogènes.
- La recherche **sur les noms de variables** est un besoin commun de la communauté scientifique des zones critiques et plus généralement de la communauté des sciences de la terre.

## Solutions apportées par la création d'un vocabulaire contrôlé de variables environnementales : [le thésaurus Theia/OZCAR](#)

- ⇒ Faciliter la découverte des données en fournissant des critères de recherche utilisant des noms de variables harmonisés
- ⇒ Améliorer la réutilisation et l'interopérabilité des données en fournissant des informations riches sur les variables permettant une interprétation non ambiguë des données par une communauté d'utilisateurs plus large que celle qui les a produites.

Theia/OZCAR thesaurus

English

×

Search

Alphabetical

Hierarchy

Groups

A B C D E F G H I K L M N O P

R S T U V W Y Z 0-9

Aboveground

Aboveground dry vegetation biomass

Aboveground herbaceous plant mass

Absolute humidity

Absorbance

Absorbed radiation

Abundance

Accumulation

Accumulation since last measurement

Accumulation since last raingauge bucket tip

Accumulation since the beginning of the year

Acetochlor

Acclonifen

Acoustic investigation variable

Acoustic velocity

Acoustic wave

Actinothermal index

Actual evapotranspiration

Actual evapotranspiration

Actual evapotranspiration of peatland

Adsorption coefficient

Aerosol variable

Air

Air absolute humidity

Air constraint

Air pressure

Air pressure at 20 meters height

Air pressure at 5 meters height

Air relative humidity

Air relative humidity at 15 meters height

Air relative humidity at 18 meters height

Air relative humidity at 2 meters height

Air relative humidity at 20 meters height

Air relative humidity at 5 meters height

Air specific humidity

Air temperature

Air temperature at 15 meters height

Vocabulary information

TITLE

DESCRIPTION

CREATOR

LICENSE

LANGUAGE

SOURCE

CREATED

LAST MODIFIED

DC:REQUIRES

TYPE

URI

Theia/OZCAR thesaurus

Thesaurus for in situ data from Environmental and Critical Zone Sciences.  
Used by Theia/OZCAR information system : <https://in-situ.theia-land.fr/>

Charly Coussot <https://orcid.org/0000-0002-0544-4802>  
Véronique Chaffard <https://orcid.org/0000-0003-2823-7117>  
Isabelle Braud <https://orcid.org/0000-0001-9155-0056>  
Sylvie Galle <https://orcid.org/0000-0002-3100-8510>

<http://creativecommons.org/licenses/by/4.0/>

<http://lexvo.org/id/iso639-3/eng>

GCMD Science Keywords: <https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords>

Monday, January 1, 2018 00:00:00

Friday, September 15, 2023 13:27:27

<http://purl.org/voc/cpm>  
<https://w3id.org/iadpt/ont>

<http://www.w3.org/2004/02/skos/core#ConceptScheme>

<https://w3id.org/ozcar-theia>

Download this vocabulary: RDF/XML Turtle

### Exemples de détails attendus :

- "Precipitation amount": intégration sur quel pas de temps ? s'agit-il de précipitations solides ou liquides (neige) ?
- "Water level": eau de surface ? Nappe phréatique ?

## Pourquoi avons-nous besoin d'un cadre pour nommer les variables ?

### Problème !!

Plus le concept de la variable est précis, plus il est difficile de trouver des relations de similitude avec des concepts d'autres vocabulaires.

➔ **Interopérabilité** 

## Pourquoi avons-nous besoin d'un cadre pour nommer les variables ?

### Problème !!

Plus le concept de la variable est précis, plus il est difficile de trouver des relations de similitude avec des concepts d'autres vocabulaires.

➡ Interopérabilité 

### Solution

Décomposer le concept de variable complexe en concepts atomiques plus simples et définir les relations de similarité entre ces concepts plus simples et les concepts d'autres vocabulaires.

➡ Interopérabilité 

## L'ontologie I-ADOPT

Décomposition des noms de variables en concepts atomiques :

- Property
- Entity: rôles [ObjectOfInterest, ContextObject, Matrix]
- Constraint

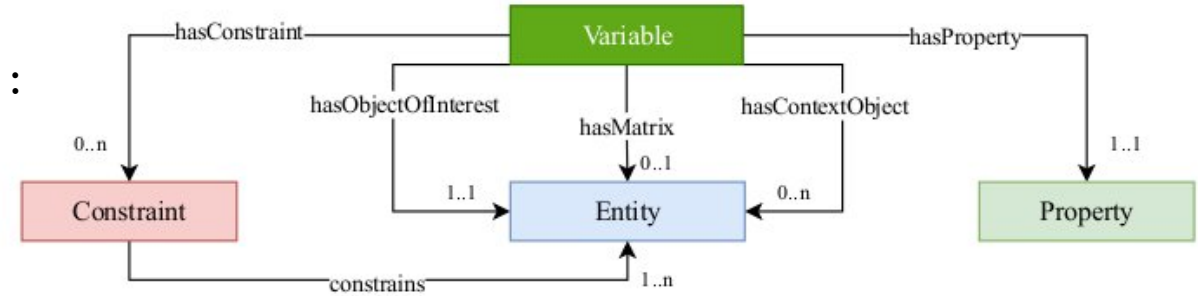
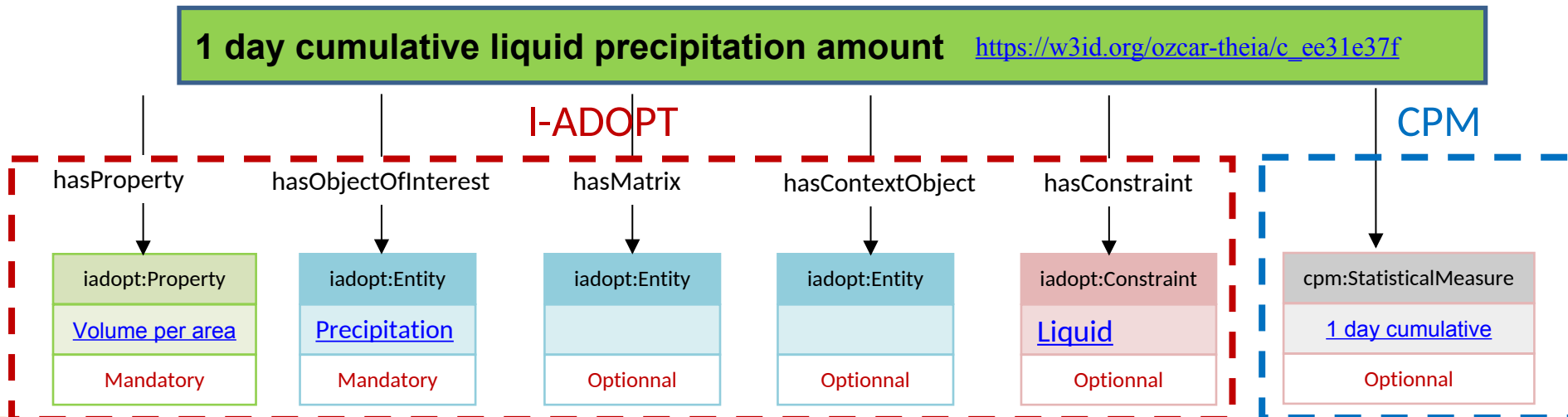


Figure 2: The I-ADOPT Framework.

*Magnana et al., S4BioDiv2021, 2021*

- Mise en œuvre facilitée par [les modèles I-ADOPT](#) fournis sur les variables quantitatives/qualitatives
- Nécessité de compléter par l'ontologie CPM (Complex Property Model) pour la notion d'agrégation temporelle et spatiale







## Ce qui fonctionne bien

- Ressources en ligne suffisantes pour maîtriser l'ontologie.
- Suffisamment générique pour modéliser toutes nos variables (atmosphère, biosphère, cryosphère, surface continentale, hydrosphère, chimie, géophysique ...)
- Enrichissement de notre vocabulaire avec de nouveaux concepts qui pourront être utilisés pour créer de nouvelles dimensions pour la découverte de données : processus, phénomènes, entité chimiques, entité environnementales (lac, rivière, ...) ...
- Combinaison avec les relations hiérarchiques SKOS pour fournir une catégorisation et améliorer la découverte de données
- Combinaison avec l'ontologie CPM pour décrire l'agrégation statistique
- Satisfait nos besoins en matière d'interopérabilité sémantique

... > Land surface variable > Soil variable > Soil geophysical variable > Acoustic investigation variable > Soil reflected acoustic wave median amplitude in all directions inside borehole

PREFERRED TERM	Soil reflected acoustic wave median amplitude in all directions inside borehole 
TYPE	Variable
BROADER CONCEPT	Acoustic investigation variable
STATISTICAL MEASURE	360° median
HASCONSTRAINT	Reflected wave
HASCONTEXTOBJECT	Borehole
HASMATRIX	Soil
HASOBJECTOFINTEREST	Acoustic wave
HASPROPERTY	Amplitude
SIMPLIFIED LABEL	Soil reflected acoustic wave amplitude
URI	<a href="https://w3id.org/ozcar-theia/c_1731d463">https://w3id.org/ozcar-theia/c_1731d463</a> 
DOWNLOAD THIS CONCEPT:	RDF/XML TURTLE JSON-LD <span>Created 12/20/22, last modified 12/20/22</span>

## Les difficultés rencontrées

- Les noms de variables sont souvent trop compliqués pour être utilisés directement pour la recherche de données. Nous avons mis en place notre propre "*simplifiedLabel*" à utiliser sur le portail web.

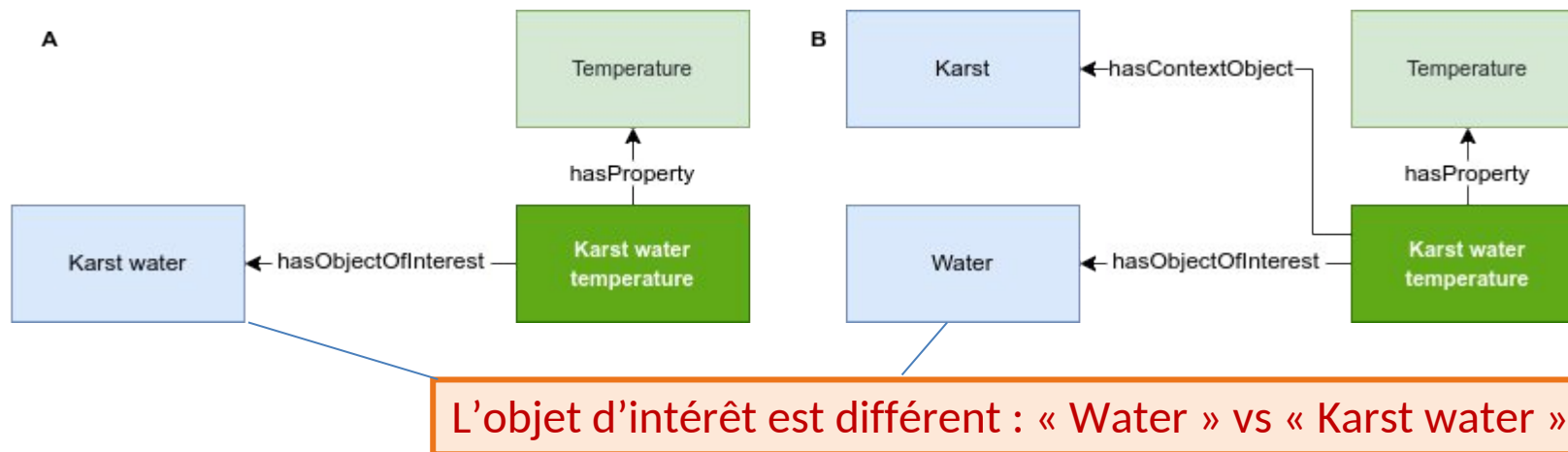
PREFERRED TERM

**1 day cumulative liquid precipitation amount** →

SIMPLIFIED LABEL

Precipitation amount

- Pour certaines variables, différentes implémentations sont possibles. Comment choisir l'une ou l'autre ? Comment déduire une relation de similitude entre des variables identiques modélisées différemment ?

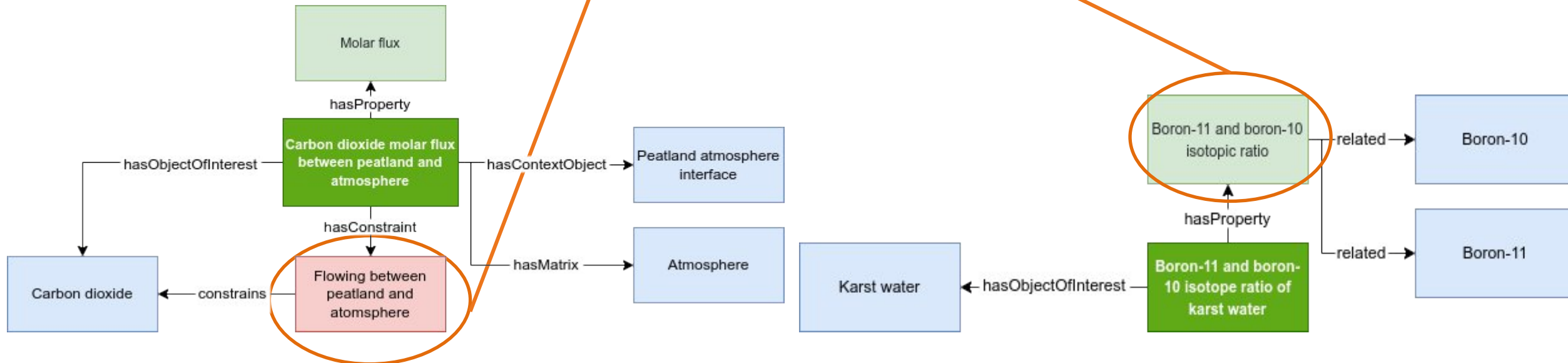




## Les difficultés rencontrées

- Difficulté à modéliser certaines variables telles qu'un flux entre deux compartiments (ex : flux de dioxyde de carbone entre la surface terrestre et l'atmosphère).
- Certains concepts doivent être très spécifiques pour décrire la variable. Ce qui peut entraîner une perte de la notion d'atomicité des concepts I-ADOPT.

Des concepts très spécifiques remettant en cause la notion initiale de concept atomique



## Les bénéfices de ce travail



- ✓ Permet une interprétation sans ambiguïté des données, et donc une meilleure réutilisation
- ✓ Favorise l'alignement des thésaurus internationaux dans le domaine.
- ✓ Favorise une meilleure interopérabilité sémantique des données au niveau national/international

Permet des études interdisciplinaires nécessitant des références croisées de données multi-sources et multi-thèmes

Permet d'ajouter des informations précises sur les variables mesurées par un jeu de données, en utilisant les champs " keyword " des standards de métadonnées qui n'incluent pas de description des variables mesurées.

The screenshot shows the eLTER data portal interface. At the top, there are navigation links: 'Discover eLTER Data', 'Visualization Time Series Data', and 'Discovery Service'. Below this is a search bar with a 'Search' button and a 'Map' button. A 'Back to search' button is also visible. The main content area displays search results for 'OZCAR-Theia variables thesaurus'. The results are organized into a table with two columns: 'GEMET - INSPIRE themes, version 1.0' and 'OZCAR-Theia variables thesaurus'. The 'OZCAR-Theia variables thesaurus' column lists four variables: 'Dissolved aluminium mass concentration per unit volume in karst water', 'Dissolved arsenic mass concentration per unit volume in karst water', 'Dissolved barium mass concentration per unit volume in karst water', and 'Dissolved hydrogenocarbonate mass concentration per unit volume in karst water'. Each variable has a magnifying glass icon next to it, indicating a search or filter function. The 'GEMET - INSPIRE themes, version 1.0' column lists 'Environmental monitoring facilities' with a magnifying glass icon.

## To learn more about the project:

Braud, I., Chaffard, V., Coussot, C., Galle, S., et al., 2020. Building the Information System of the French Critical Zone Observatories network: Theia/OZCAR-IS, Hydrological Sciences Journal, special issue “Data: opportunities and barriers”,

<https://doi.org/10.1080/02626667.2020.1764568> .

Coussot et al., Putting in practice the I-ADOPT framework for the naming of environmental variables from continental surfaces, in preparation

## To access the portal, the thesaurus and the project Github

Data portal : <https://in-situ.theia-land.fr/>

Thesaurus: <https://w3id.org/ozcar-theia/>

Cataloguing CSW webservice: <https://in-situ.theia-land.fr/geonetwork/srv/eng/csw?service=CSW&version=2.0.2&request=GetCapabilities>

GitHub : <https://github.com/theia-ozcar-is>

## Contacts:

[charly.coussot@univ-grenoble-alpes.fr](mailto:charly.coussot@univ-grenoble-alpes.fr)

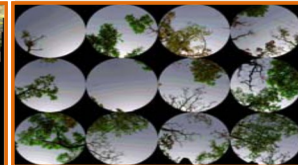
[Veronique.Chaffard@univ-grenoble-alpes.fr](mailto:Veronique.Chaffard@univ-grenoble-alpes.fr)

[Isabelle.braud@inrae.fr](mailto:Isabelle.braud@inrae.fr)

[Sylvie.galle@ird.fr](mailto:Sylvie.galle@ird.fr)

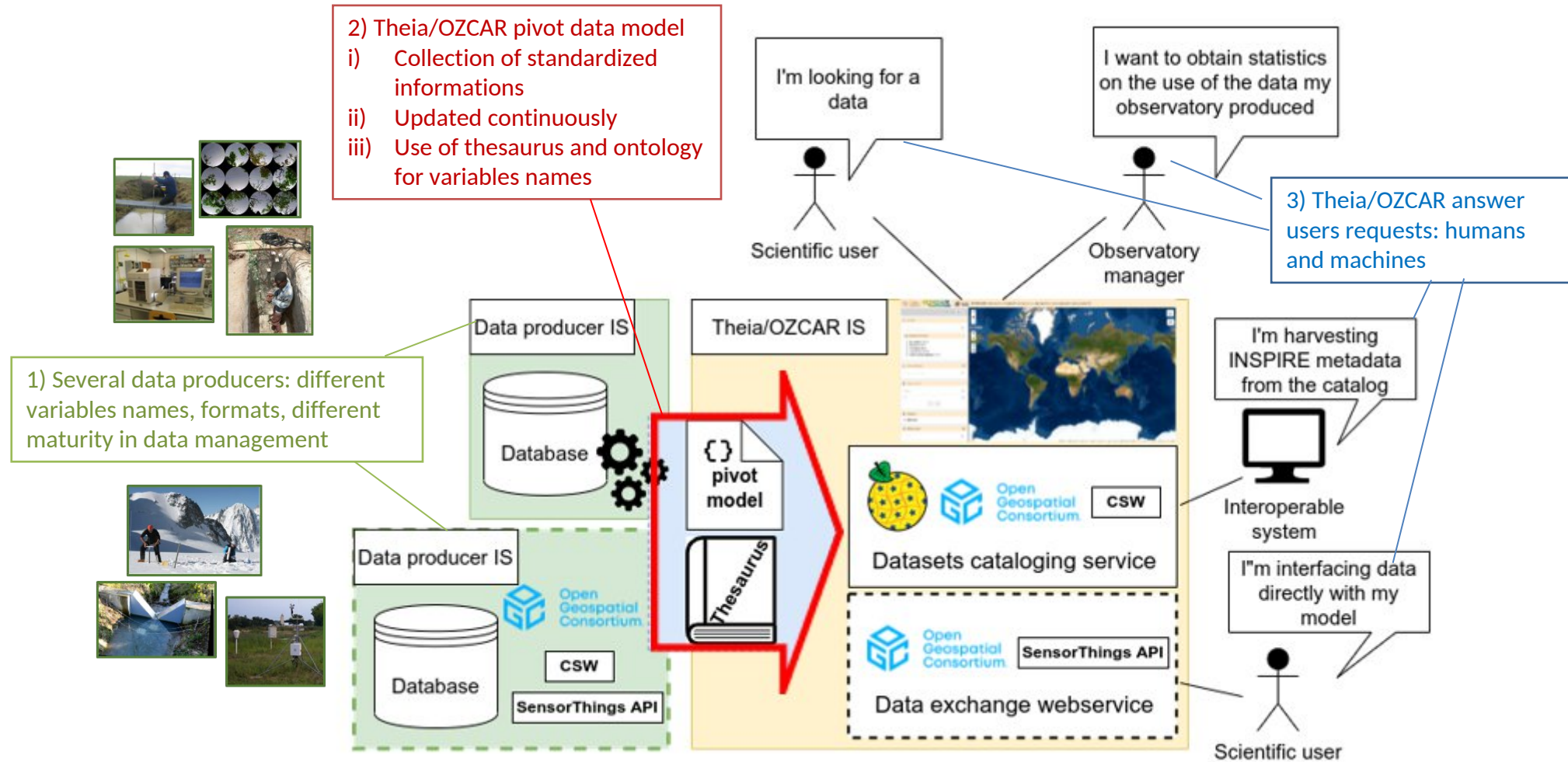
**Acknowledgements:** The project contributes to the Theia, continental surface data pole and the Data Terra Research Infrastructure. It receives supports from the OZCAR Research Infrastructure, IRD, CNRS and was partly funded by the FAIRTois ANR Project (grant ANR-19-DATA).

# Merci pour votre attention



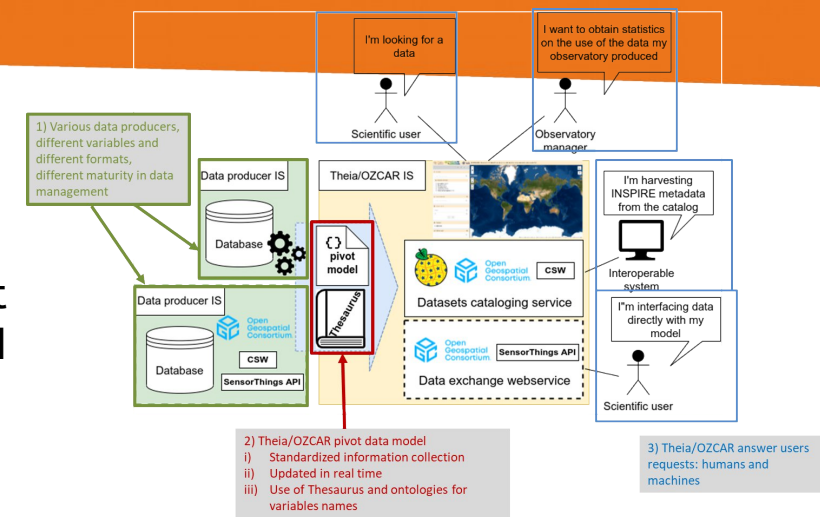


## Data fluxes and services between data producers and users

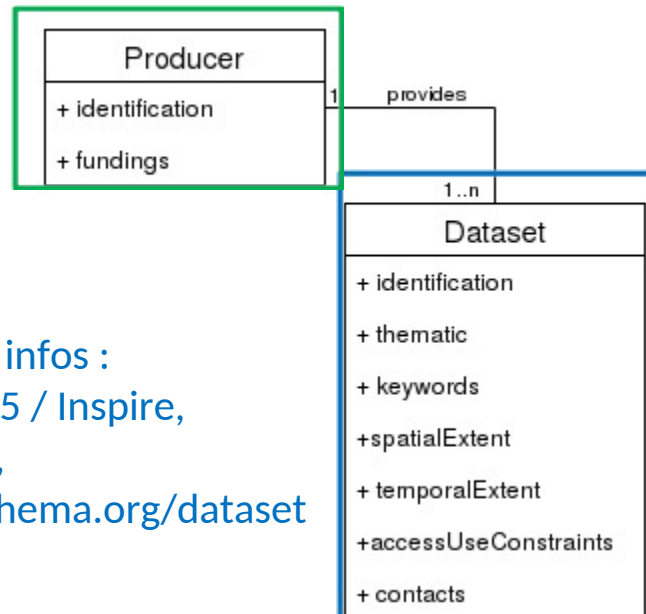


## Pivot data model

- **Pivot data model** to harmonize data description, get the required information for the faceted search and set up data exchange web services
- Based on the mapping of different standards

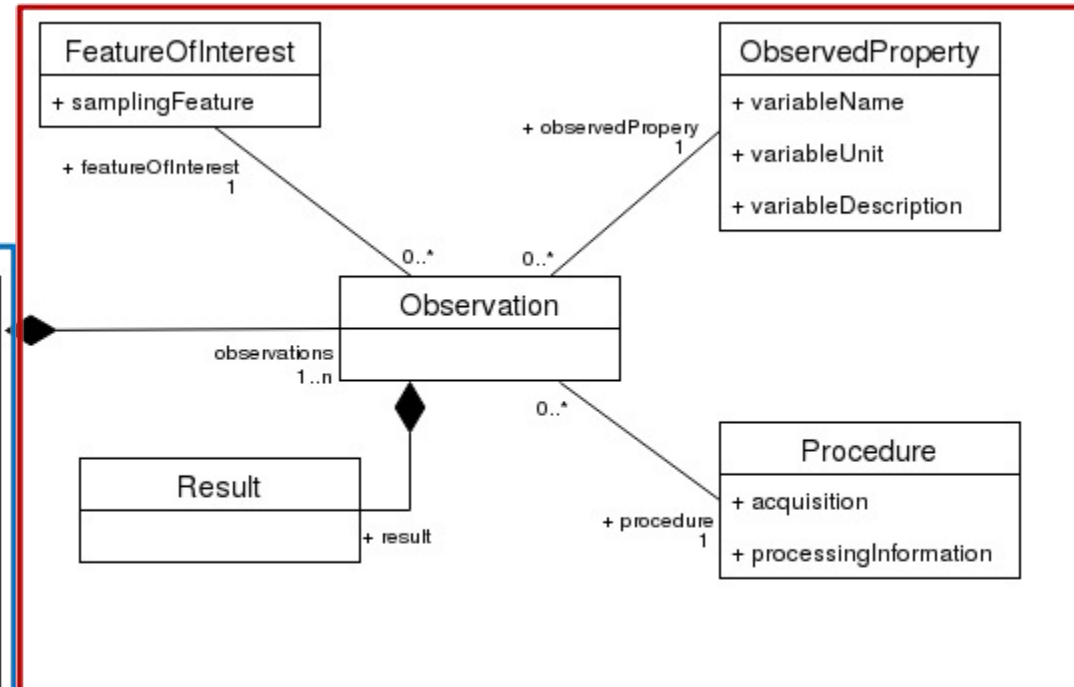


### Producers info (DataCite, scanR API)



Datasets infos :  
ISO 19115 / Inspire,  
DataCite,  
DCAT, [schema.org/dataset](https://schema.org/dataset)

### Observations infos: O&M



<https://github.com/theia-ozcar-is/data-model-documentation>



A FAIR thesaurus FAIR : <https://w3id.org/ozcar-theia>

DOI : 10.17178/67b5a1d5-8c8c-4a94-a646-1cca1d0adf79

Variables

Objects of interest

Theia/OZCAR thesaurus

English

Search

Alphabetical

Hierarchy

Groups

Constraint

Instrument

Method

Observable property

Phenomenon

Physical entity

Process

Time

Variable

Vocabulary information

TITLE

Theia/OZCAR thesaurus

DESCRIPTION

Thesaurus for in situ data from Environmental and Critical Zone Sciences.  
Used by Theia/OZCAR information system : <https://in-situ.theia-land.fr/>

CREATOR

Charly Cousot <https://orcid.org/0000-0002-0544-4802>  
Véronique Chaffard <https://orcid.org/0000-0003-2823-7117>  
Isabelle Braud <https://orcid.org/0000-0001-9155-0056>  
Sylvie Galle <https://orcid.org/0000-0002-3100-8510>

LICENSE

<http://creativecommons.org/licenses/by/4.0/>

LANGUAGE

<http://lexvo.org/id/iso639-3/eng>

SOURCE

GCMD Science Keywords: <https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords>

CREATED

Monday, January 1, 2018 00:00:00

LAST MODIFIED

Friday, July 1, 2022 13:45:37

DC:REQUIRES

<http://purl.org/voc/cpm>  
<https://w3id.org/iadopt/ont>

TYPE

<http://www.w3.org/2004/02/skos/core#ConceptScheme>

OZCAR Theia in-situ thesaurus

Alphabetical

Hierarchy

Groups

Variable

Atmosphere variable

Biosphere variable

Cryosphere variable

Land surface variable

Terrestrial hydrosphere variable

Groundwater hydrology

Karst hydrology

Surface water hydrology

Surface water chemistry

Surface water microbiology

Surface water physic variable

Pond turbidity

River discharge

Surface water conductivity

Surface water pH

Surface water suspended sediment concentration

Surface water temperature

Water level

OZCAR Theia in-situ thesaurus

Alphabetical

Hierarchy

Groups

Physical entity

Chemical entity

Environmental entity

Atmosphere

Biosphere

Cryosphere

Hydrosphere

Groundwater

Karst water

Water table

Surface water

Lake

Pond

River

Spring

Water

Cloud

Dew

Raindrop

Land surface

Grain

Rock

Sediment

Soil

Topography

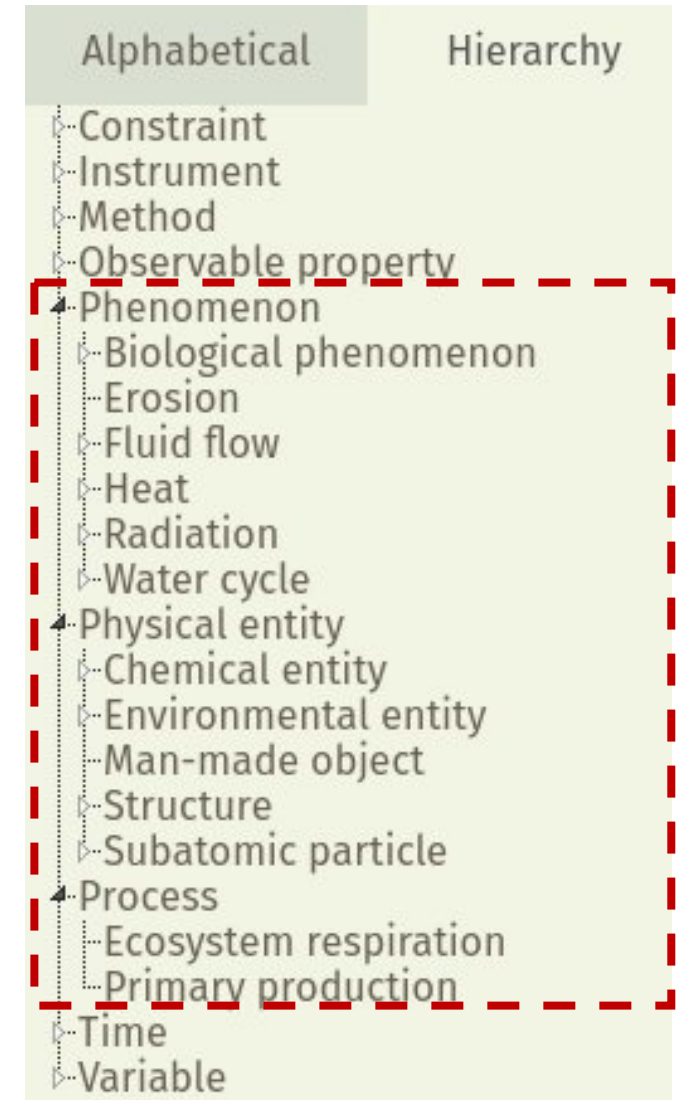
## Benefits

Description of variables with rich and formal semantics (ad-hoc ontology)

✓ Enriched our thesaurus with new concepts

ObjectOfInterest : process, phenomenon, chemical entity,  
environmental entity (lake, river, ...), ...

-> which would allow us to offer new search dimensions on our portal (in addition to the observed variable)



## Benefits

- ✓ Promotes unambiguous interpretation of data and therefore better reuse

### Variable

PREFERRED TERM	1 day cumulative liquid precipitation amount
-----	
TYPE	Variable
BROADER CONCEPT	Precipitation amount
STATISTICAL MEASURE	1 day cumulative
HASCONSTRAINT	Liquid
HASOBJECTOFINTEREST	Precipitation
HASPROPERTY	Volume
SIMPLIFIED LABEL	Precipitation amount
URI	<a href="https://w3id.org/ozcar-theia/c_ee31e37f">https://w3id.org/ozcar-theia/c_ee31e37f</a>
DOWNLOAD THIS CONCEPT:	RDF/XML TURTLE JSON-LD

### ObjetOfInterest

PREFERRED TERM	Precipitation
-----	
TYPE	Entity
DEFINITION	[Wikipedia] In meteorology, precipitation is any product of the condensation of atmospheric water vapor that falls under gravitational pull from clouds. The main forms of precipitation include drizzling, rain, sleet, snow, ice pellets, graupel and hail. Precipitation occurs when a portion of the atmosphere becomes saturated with water vapor (reaching 100% relative humidity), so that the water condenses and "precipitates" or falls.
BROADER CONCEPT	Water cycle
URI	<a href="https://w3id.org/ozcar-theia/c_2b48133e">https://w3id.org/ozcar-theia/c_2b48133e</a>
DOWNLOAD THIS CONCEPT:	RDF/XML TURTLE JSON-LD