

## Theia/OZCAR pivot data model

### Metadata file in JSON format

#### Version 1.2

##### Changelog:

« *producer.contacts.\** » : Removal of the constraint of having only one Project leader.

« *datasets.\*.observations.\*.featureOfInterest.samplingFeature.name* » : Change from recommended to mandatory field. The name chosen must be persistent over time

« *datasets.\*.metadata.portalSearchCriteria* » : Change from mandatory to optional.

« *datasets.\*.metadata.portalSearchCriteria.geologies* » : Change from mandatory to optional/deprecated given that the enrichment of data according to geology is carried out automatically at the level of each observation.

« *datasets.\*.metadata.portalSearchCriteria.climates* » : Change from mandatory to optional/deprecated given that the enrichment of data according to climate is carried out automatically at the level of each observation.

« *datasets.\*.observations.\*.procedure.dataProduction.sensors.\*.model* » : Change from mandatory to recommended.

« *datasets.\*.observations.\*.procedure.dataProduction.sensors.\*. manufacturer* » : Change from mandatory to recommended.

« *datasets.\*.observations.\*.procedure.dataProduction.sensors.\*. sensorType* » : Change from optional to mandatory.

« *datasets.\*.observations.\*. observedProperty. keywords* » : Addition - optional field.

« *datasets.\*.observations.\*. observedProperty. gcmdKeywords* » : deprecated as it is replaced by the more generic keywords field.

« *datasets.\*.observations.\*.observedProperty.theiaCategories* »: It is no longer mandatory to fill in the URIs of the last level of variables categories. The URIs of variables categories should identify the compartment of the critical zone in which the observation was made. Alternatively, it is now possible to reference an URI of variable rather than an URI of category if this term of type "Variable" already exists in the thesaurus and describes the producer variable precisely enough. In this case, the JSON table will contain one and only one URI representing the Theia/OZCAR variable with which to associate the producer variable.

Translated with [www.DeepL.com/Translator](http://www.DeepL.com/Translator) (free version)

## Foreword :

The information flux to be set up between data producers and the Theia/OZCAR IS is based on a "pivot" data-model bringing together a set of metadata to be exchanged in order to promote data discovery, sharing and reuse. It will provide the functionalities of the Theia/OZCAR data access web portal and will allow the implementation of different services such as metadata cataloguing service or service for obtaining DOI for datasets.

The analysis of the metadata to be provided focused initially on time series data for point/trajectory based observations. The format will be adapted to other types of data.

**The purpose of this document is to describe the content of the pivot data model to be implemented in JSON format.** This JSON text file will be produced by each data producer. The Theia/OZCAR portal will then analyse the information contained in this file. This information includes information on the data producer and the datasets including the variables that compose them. A dataset groups geographically referenced data on one or more variables observed over a time period. The dataset is described by metadata describing its content and the variables measured.

**JSON** (JavaScript Object Notation) is a lightweight data exchange format. It is easy for humans to read or write. It can be analysed or generated by machines. It is based on a subset of the JavaScript [programming language \(JavaScript Programming Language, Standard ECMA-262 3rd Edition - December 1999\)](#). JSON is a text format completely independent of any language. JSON is based on the following structures:

An **object**, which is a set of non-ordered key/value pairs. An object begins with { (left brace) and ends with } (right brace). Each name is followed by : (colon) and the key/value pairs are separated by , (comma). The set of key/value pairs are the members of the object.

An **array** is a collection of ordered values. A table begins with [ (left square bracket) and ends with ] (right square bracket). The values are separated by , (comma). The collection of ordered values are the elements of the table.

A **value** can be either a string of characters in quotation marks, a number, `true` or `false` or `null`, an **object** or an **array**. These structures can be nested.

Geographic objects are described according to the [GeoJSON](#) standard, which allows a wide variety of geometric structures to be described.

## Choice of language for metadata:

The choice of language (English or French) describing the metadata is left to the initiative of the data provider according to what exists in its information system. If there is no existing one, English is preferred.

On the other hand, **the possible values of the enumerations (see: annex) must be provided in English.**

If both languages exist, it is possible to provide us with both; file names indicating the language used with a 2 digit code (e.g. CATC\_fr.json)

## **Nomenclature of the json file name:**

The file name must comply with the following convention:

XXXX\_yy.json

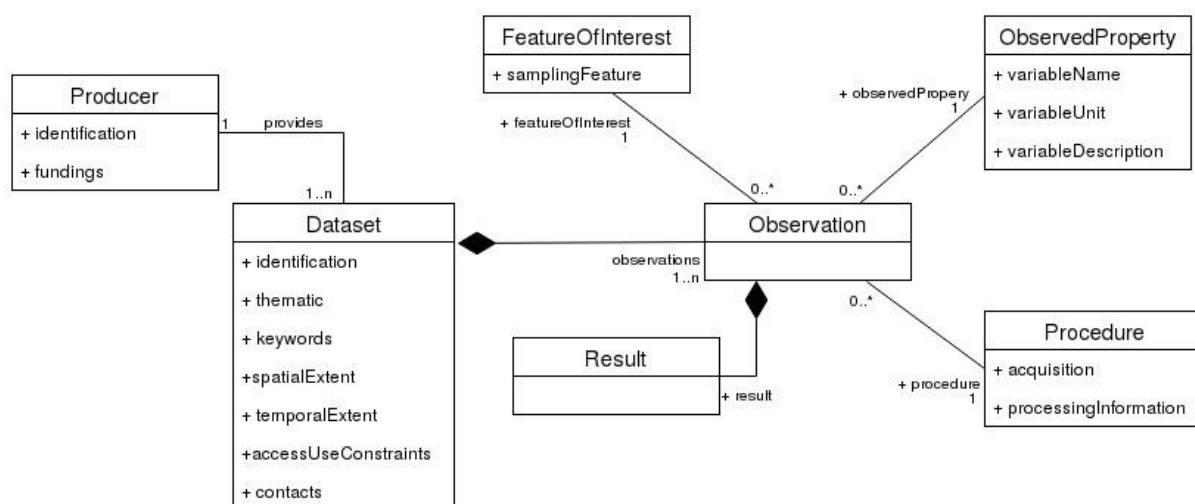
XXXX: The Theia/OZCAR identifier of the data provider. The list of identifiers for each of the observatories is available in the appendix.

yy : the two-letter code of the language used in the file. "en" for English or "fr" for French. In order to implement interoperability with European infrastructures, an English language file is required.

## **Encoding the characters to be used in the files:**

The encoding to use is UTF-8.

## **Simplified class model of the pivot format**



The information describing a dataset (Dataset class) is derived from ISO 19115/Inspire.

The concepts describing an observation are derived from the OGC O&M standard: Observation classes, FeatureOfInterest, ObservedProperty, Procedure, Result

## Mandatory / recommended / optional metadata:

The mandatory or non-mandatory nature of the metadata is indicated in the document with the following code:

- **M** (mandatory) = **mandatory information**
- **R** (recommended) = **recommended information**
- **O** (optional) = **optional information**

A JSON object can consist of a set of JSON objects. An optional JSON **O** object can be made up of mandatory JSON **M** objects. In this case, if the optional JSON object is implemented it will be made up of its mandatory members.

## Management of non-values for optional json fields

For optional string fields that the producer does not want to fill in, and in order to comply with our defined json schemas, it is expected that these **fields are absent from the json file**. Values equal to an empty string (for the string type) or null are not allowed.

Ex : **O** email

To do : Absence of the field

Not to do : "email": " " or "email": null

## Non-value code for units of measurement for variables

For **variables without unit** (ex : pH), the value of the field "**unit**" must be filled with the **code "N/A"**.

Ex : "observedProperty": {

    "name": "pH",

    "unit": "N/A",

    ....

}

## Detailed description of the objects in the json file:

In this document, each name of a key/value pair is indicated on a grey background (**name**).

At the highest level of the file are defined 3 JSON objects with the following names:

- **M producer**: is the metadata sheet relating to the data provider publishing the file
- **M datasets**: is the observatory's collection of data sets. This collection is represented by a JSON table whose elements represent each supplier's dataset.

**M version** : is the character string corresponding to the version number of the pivot format. This first version is "1.0".

### 1. The JSON **producer** object

The **producer** object is composed of the following members:

- **M producerId**: Identifier of the data provider. The observatories' identifiers are provided in the appendix.
- **M name** : key/value pair whose value is the name of the observatory. *Example: "AMMA-CATCH"*. Free text field.
- **M title**: key/value pair whose value is the observatory's title. *Example: "AMMA-CATCH: a hydrological, meteorological and ecological observatory on West Africa"*. Free text field.
- **M description** : key/value pair whose value is a summary of the data provider's scientific activities. Free text field required. *Example: "The Draix-Bléone Observatory was created in 1983 to study hydrological and erosive processes in the mountains. It is located in the Southern Alps, on a very erodible substrate of black marl, and characterized by a mountainous and Mediterranean climate. About ten catchment areas have been equipped for measuring solid and liquid flows. These basins cover areas ranging from 0.1 ha to several hundred km2, with varied plant cover, lithologies and topographies The Draix-Bleone Scientific Interest Group, created in 2002, currently brings together 18 laboratories with skills in hydrology, hydraulics, geomorphology, ecology, geochemistry, engineering and geography. »*

- **R objectives**: key/value pair whose value is a summary of the scientific objectives of the data provider. Example: *"The objective of the Draix-Bléone observatory is to study hydrological and erosive processes in the mountains. »*
  - **R measuredVariables**: key/value pair whose value is a summary of the variables measured by the data provider. Example: *Available data include rainfall and climate data (temperature, humidity, radiation), flow chronicles at the outlet of the monitored basins, suspended solids flows and coarse material deposition for each flood, soil water status monitoring, numerical terrain model and vegetation cover.*
  - **O email** : key/value pair whose value is the string of characters representing the email of the data provider.
  - **M contacts**: collection of contacts for the data producer represented in the form of a JSON table. **Contacts must be natural persons**: two roles are possible. The **Data Manager** is the person who acts as a data manager. The **Project Leader** is the scientific manager of the observatory. These roles are useful for creating DOIs and publishing INSPIRE metadata sheets. At least one contact with the **Project leader** role is required. The elements of the **contact** collection representing natural persons are objects with the following members:
    - o **M firstName**: the contact's first name. Free character string. This element is mandatory.
    - o **M lastName**: the name of the contact. Free character string. This element is mandatory.
    - o **M email**: email address of the contact. String of characters. This element is mandatory.
    - o **M role**: The role of the contact for the data set. This field is required. The field can be one of the following values: "Project leader", "Data manager".
    - o **O orcid**: the orcid identifier of the person if it exists.
    - o **O organisation**: the organisation of the physical contact's affiliation. A JSON object composed of the following members:
      - **M role**: The role of the organization, for the moment only one role is possible "Research group".
      - **M iso3166**: the two-letter code of the country of the organization. See [http://kirste.userpage.fu-berlin.de/diverse/doc/ISO\\_3166.html](http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html)
      - **R name**: the name of the organization to which it belongs.
      - **R acronym**: the acronym for the name of the organization to which it belongs.
      - **R idScanR** : id ScanR of the organization (SIREN or id RNSR)
- A member of the organisation object must be composed of an acronym or a name or an idScanR**
- **M fundings**: a JSON table representing the list of funders of the data provider (operation and payroll). This field is required. It is useful for DOI reporting as well as for producing statistics on financial guardianship. Each element of the JSON **fundings** table is a JSON object composed of the following members:
    - o **M type**: the type of funder respecting the list in the appendix. The field can take one of the values from the "EnumFundingType" enumeration. A funder can be a research institute (e.g. CNRS, IRD, BRGM, IGN...), a university, an OSU is of the "French Federative structure" type. The UMRs are of the "Research unit" type.
    - o **M iso3166**: the two-letter code of the funder's country. See [http://kirste.userpage.fu-berlin.de/diverse/doc/ISO\\_3166.html](http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html)
    - o **R acronym**: the funder's acronym
    - o **R name**: the name of the funder
    - o **R idScanR** : id ScanR of the organization (SIREN or id RNSR). Visible on <https://scanr.enseignementsup-recherche.gouv.fr/>

A member of the **Funding** object must be composed of an **acronym** or a **name** or an **idScanR**

- **O onlineResource**: The different online resources related to the data producer. This JSON object is composed of the following members:
  - o **R urlDownload**: This metadata element provides a link to the data producer's data access portal. Text field in the form of a valid url.  
Ex: <http://sevnol.ohmcy.fr/Sevnol2>
  - o **R urlInfo**: This metadata element provides a link to the data producer's (observatory's) website. Text field in the form of a valid url.  
Ex: <http://www.ohmcy.fr>
  - o **R doi**: The identifier must uniquely reference the producer (doi of the entire observatory). Text field in the form of a valid doi identifier. Ex: 10.17178/AMMA-CATCH.all
  - o **R webservices**: JSON table listing the URLs of the different web services allowing the data producer to be interoperable. Each element of the JSON **webservices** array is a JSON Object composed of two key/value pairs. The first key/value pair with key=**description** is a free text field describing the webservice. The second key/value pair with key= **url** is a free text field providing the valid web service url. This field will allow to distribute the webservices set up on the data producer's information. An ISO19115/INSPIRE sheet published in a CSW catalogue for the data producer will be described in the **description** field by the url of the *getRecordById* whose id corresponds to the data provider's sheet. Ex: <https://sig.oreme.org:8443/geonetwork/srv/eng/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetRecordById&ID=8dbdb4d7-161a-45e3-b3c3-51e6058f6e8a>. A CSW catalogue of the data sets of the data producer will be described in the **description** field by the *getCapabilities* url. Ex: <https://sig.oreme.org:8443/geonetwork/srv/eng/csw?SERVICE=CSW&REQUEST=GetCapabilities&VERSION=2.0.2>

## 2. The JSON **datasets** object:

The **datasets** object is the collection of the datasets of the data provider. The JSON **datasets** object is represented as a JSON array. Each value in the JSON array is a JSON object describing a dataset. Each JSON object representing a data set has the following 3 members:

- **M datasetId**: a character string composed of a code formed by the first 4 letters of the name of the data provider according to the list provided in Annex 1, followed by a character string "\_DAT\_" and a **permanent identifier of the data set**. **It is mandatory that the identifier chosen for the dataset does not change over time**. It can be an identifier of a database table or a code identifying the dataset  
Ex: AGRH\_DAT\_1
- **M metadata**: a JSON object containing the metadata describing the dataset
- **M observations**: a JSON object collecting the observations associated with the dataset.

### a) The **metadata** object:

This JSON object is a member of each of the JSON objects in the **datasets** collection. It itself contains the following members:



- **M title:** the title of the dataset. It is recommended to include an indication of the geographical area covered, especially if the same type of scientific data recording is proposed in several territories.

It is recommended not to include responsible parties, i.e. organisations with any responsibility (manager, owner, etc.).

It is recommended not to indicate only the acronym of the resource (unless it is explained in the description field).

Mandatory free text field that will be useful for creating INSPIRE compliant metadata records.

*Example: Groundwater level in rural wells, Oueme meso-site, Benin.*

- **M description:** a summary of the dataset. The summary must describe the resource in a way that is understandable to the user. For a producer, it is in particular a question of defining as well as possible the information or phenomenon represented in the data. It will therefore contain elements of definition, but also possibly a summary indication of the area covered or, if applicable, information on the particularities of the version of the dataset.

Mandatory free text field that will be useful for creating INSPIRE compliant metadata records.

*Example: Groundwater level measurements and manual conductivity measurements for some wells. Hourly measurements for equipped wells. The groundwater level is expressed from the top of the well.*

- **O objective:** Summary of the scientific objectives of the dataset. This element is optional. Free text field.
- **O onlineResource:** The different online resources of the dataset. This JSON object is composed of the following members:

- o **R urlDownload:** This metadata element provides a link to a web page for downloading the dataset. This is recommended if such a possibility exists. Text field in the form of a valid url. It will redirect the user during a download action.

*Example: <ftp://ftp.ifremer.fr/ifremer/argo/>*

- o **R urlInfo:** This metadata element provides a link to a web page describing the dataset. Text field in the form of a valid url.

*Example: <http://www.seanoe.org/data/00311/42182/>*

- o **R doi:** The identifier must uniquely reference the dataset. Text field in the form of a valid doi identifier. It will enable the citation of the dataset.

*Example: 10.13155/29825*

- o **R webservices:** JSON table listing the URLs of the different web services allowing the interoperability of the dataset. Each element of the JSON **webservices** array is a JSON Object composed of two key/value pairs. The first key/value pair with key=description is a free text field describing the webservice. The second key/value pair with key= url is a free text field providing the valid web service url. This field will allow to distribute the webservices of the dataset. These could be used by the application in future versions. An ISO19115/INSPIRE sheet published in a CSW catalogue for the dataset will be described in the **description** field by the url of the *getRecordById* whose id corresponds to the dataset sheet. Ex:

<https://sig.oreme.org:8443/geonetwork/srv/eng/csw?SERVICE=CSW&VERSION=2.0.2&REQUEST=GetRecordById&ID=131142c1-cb88-4510-9f52-e0fe5f3c18be>

- **M datasetLineage:** describes the genealogy of a dataset, i.e. the history of the dataset and, if known, its life cycle, from the acquisition and entry of information to its compilation with other datasets and variations of its current form. It is a question of providing a literal and concise description either of the history of the dataset or of the means, procedures or computer processing implemented when the dataset was acquired. Genealogy refers to the processing history and/or general quality of the geographic data set. Where appropriate, it may include information indicating whether the data set has been validated or subjected to quality control, whether it is the official version (in case there are several versions) and whether it has



legal value. This element is a free character string and is mandatory for creating INSPIRE compliant metadata records.

- **M contacts:** collection of contacts for the data set represented in the form of a JSON table. A member of the contact collection can be either a natural person or an organization.

**In the case of a natural person:** four roles are possible. Two "technical" roles, Data Manager or Data Collector. The Data Manager is the person who performs the role of data manager. The Data Collector is the person who collects the data. If a contact with the Data Manager role is specified, this will be the contact point for the dataset. Two "scientific" roles are possible: Principal Investigator (scientific referrer for the data set), or Project Member (scientist associated with the creation of the data set). These roles are useful for creating DOIs and publishing INSPIRE metadata records. At least one contact with the Principal Investigator role is required, all other roles are optional. The elements of the **contact** collection representing natural persons are objects with the following members:

- o **M firstName:** the contact's first name. Free character string. This element is mandatory.
- o **M lastName:** the name of the contact. Free character string. This element is mandatory.
- o **M email:** email address of the contact. String of characters. This element is mandatory.
- o **M role:** The role of the contact for the data set. This field is required. The field can be one of the following values: "Principal investigator", "Project member", "Data manager", "Data collector". The role "Project leader" is defined for the Producer object only.
- o **O orcid:** the orcid identifier of the person if it exists.
- o **O organisation:** the organization of the physical contact's affiliation. A JSON object composed of the following members:
  - **R name:** the name of the organization to which it belongs.
  - **R acronym:** the acronym for the name of the organisation to which it belongs.
  - **M role:** role of the organization with the value "Research group".
  - **M iso3166:** the two-letter code of the country of the organization. See [http://kirste.userpage.fu-berlin.de/diverse/doc/ISO\\_3166.html](http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html)
  - **R idScanR:** id ScanR of the organization (SIREN or id RNSR). Visible on <https://scanr.enseignementsup-recherche.gouv.fr/>

**In the case of an organization:** only one role is possible, "Research group". The elements of the contact collection representing organizations are elements represented by JSON objects composed of the following members:

- o **R name:** the name of the organization to which it belongs.
- o **R acronym:** the acronym for the name of the organization to which it belongs.
- o **M role:** role of the organization with the value "Research group".
- o **M iso3166:** the two-letter code of the country of the organization. See [http://kirste.userpage.fu-berlin.de/diverse/doc/ISO\\_3166.html](http://kirste.userpage.fu-berlin.de/diverse/doc/ISO_3166.html)
- o **R idScanR:** id ScanR of the organisation (SIREN or id RNSR). Visible on <https://scanr.enseignementsup-recherche.gouv.fr/recherche/structures>

**A member of the organisation object must be composed of an acronym or a name or an idScanR**

- **M dataConstraint:** This metadata element defines access and use constraints, i.e., the conditions applicable to public access limitations and access and use of spatial data sets. The **dataConstraint** object is composed of the following members:

- **M AccessUseConstraint**: a free text field describing the conditions of use. If the dataset is subject to conditions of use (e. g. use under condition of quotation via thank you phrase), they must be indicated in this field. If no conditions apply, the field will be set to "No conditions apply", if the conditions are unknown, the field will be set to "Unknown conditions". If the metadata elements are described in English, the field can take "[no conditions to access and use](#)" or the value "[conditions to access and use unknown](#)". This field is required.
- **O urlDataPolicy**: provides a link to the data policy url. Text field in the form of a valid url.
- **O license**: JSON object representing the license to which the dataset is associated. The JSON object is composed of the following members:
  - **M title**: the name of the license. This field is mandatory if a license has been subscribed for the dataset.  
Ex: CC BY 4.0
  - **M url**: URL of the dataset license. This field is mandatory if a license has been subscribed for the dataset  
Ex: <https://creativecommons.org/licenses/by/4.0/>
- **O embargo**: an embargo period can be defined to limit access to the recent data. This embargo period is the duration in days, since the current date, for which access to the data is restricted to a limited number of users. Thus, for an embargo period of 365 days, if a request is made on the data set as of 01/03/2018, the data measured on 01/05/2017 will not be available, while the data measured on 28/02/2017 will be available. The JSON **embargo** object consists of two members:
  - **M duration**: the duration in days of the embargo on the dataset.
  - **M privilegedUsers**: JSON array of text field. Each field is the email of a privileged user authorized to bypass the embargo on the dataset.
- **O portalSearchCriteria**: This field allows to filter the datasets on the portal according to the following criteria :
  - **O climates**: **deprecated**. A collection of climates that can be associated with the dataset described in the form of a JSON table. The list of possible climates is available in the appendix.
  - **O geologies**: **deprecated**. A collection of geologies that can be associated with the dataset described in the form of a JSON table. The list of possible geologies is available in the appendix.

The **climates** et **geologies** fields are **deprecated** given that the enrichment of data according to climate and geology is carried out automatically at the level of each observation.

- **M topicCategories**: Thematic category of the dataset from the ISO 19115/Inspire standard. Value: JSON table of text fields according to the list of INSPIRE thematic categories provided in the appendix. At least one value from the thematic category must be provided. This field will allow the creation of INSPIRE metadata sheets.
- **M inspireTheme**: Theme of the data set required by INSPIRE. Value: text in accordance with the list of the different INSPIRE themes provided in the appendix. This field will allow the creation of INSPIRE metadata sheets. A mapping between **inspireTheme** and **topicCategories** fields is suggested in [Technical guideline for metadata entry](#) but is not mandatory.
- **O keywords**: JSON table of JSON objects. Each of the JSON objects representing a keyword. The JSON Object representing a keyword is composed of 2 key/value pairs. The first pair with key = "keyword" is a free text field describing the keyword, and the second pair with key= "uri" is a free text field describing the URI of the keyword in a published thesaurus.

- **O documents**: collection of documents associated with the dataset (publications or manuals) represented by a JSON table. Documents can be of two types: "Publication" or "Manual". The **document** object is a JSON array. Each of the elements in this table is a JSON object with two members. A "type" member can take the value "Publication" or "Manual", a "url" member representing the url to which the document is accessible. This optional element is used to document the reference publications associated with the dataset, the protocols for acquiring the dataset, etc.
- **M spatialExtent**: The extent of the game in the geographical space, expressed as one or more boundary rectangles (BBOX). This delimitation rectangle is defined by the east and west longitudes and the south and north latitudes in decimal degrees, with an accuracy of at least two decimal places. These coordinates are expressed according to the coordinate reference frame WGS84 (EPSG:4326). If several rights-of-way are indicated, they must not overlap. This field will allow the creation of INSPIRE compliant metadata sheets and document the hold of data sets. **spatialExtent** is a GeoJSON object and the following members are used to describe it.
  - o **M** The constant "type": "Feature" is assigned in order to match the object to the GeoJSON schema.
  - o **M** The constant "properties": {} is assigned in order to match the object to the GeoJSON schema.
  - o **M geometry** describes the spatial scope of the dataset. The **geometry** object is composed of the following members.
    - **M type**: the type of the geometry of the right-of-way. The two types of possible rights-of-way are "Polygon" and "MultiPolygon" in the case where several rights-of-way are required for the dataset.
    - **M coordinates**: for a polygon type, the **coordinates** member must be a collection of position collections. Each position is described by the coordinate doublet[longitude,latitude]. The coordinates are expressed in decimal degrees. The rectangular right-of-way will have 5 positions, the fifth identical to the first. The polygon coordinates are defined in the collection in a counter-clockwise direction. For a MultiPolygon type, the **coordinate** member is a Polygon collection.

## b) The object **observations**

**In the pivot format, an observation contains the time series of a given variable at a given location.** The **observations** object is the collection of observations associated with the dataset. It is described by a JSON table. According to the O&M (Observation & Measurement) standard, an "observation" can be defined as the action whose result is the value at a given time of a property (**observedProperty**) of an object of interest (**featureOfInterest**) obtained through a procedure (**procedure**).

Each element of the **observations** collection is a JSON object describing an observation composed of the following elements:

- **M observationId**: a character string composed of a code formed by the first 4 letters of the name of the data provider according to the list provided in Annex 1, followed by the character string "\_OBS\_", followed by a unique identifier of the observation. The unique identifier of the observation in the dataset can be defined incrementally and **unlike the identifier of the dataset**, it is not necessary to persist it.

Ex: AGRH\_OBS\_1

- **M observedProperty**: identifies the variable describing the observed phenomenon. This variable is related to the object of interest of the observation (**featureOfInterest**). The **observedProperty** object is a JSON object composed of the following fields:
  - **M name**: string of characters describing the name of the variable as it is called in the data provider's information system; Theia/OZCAR IS is responsible for making the associations with the Theia/OZCAR thesaurus.
  - **M unit**: character string describing the unit of measurement of the variable
  - **O description**: string proposing a description of the variable. This description may facilitate the association of the variable with the Theia/OZCAR thesaurus of variables and facilitate the understanding of the measurement performed.
  - **O gcmdKeywords**: **deprecated**. Collection of JSON objects each describing a field of the GCMD taxonomy for variable names (see: Global Change Master Directory controlled vocabularies: [Earth Science and Earth Science Services](#)). Each of the associations with the GCMD taxonomy carried out by the Theia/OZCAR team during the work phase on variable vocabulary will enrich this field. Each object in the collection is composed of the following members:
    - **M category**: the category of the GCMD field
    - **M topic**: the topic of the GCMD field
    - **O term**: the term of the GCMD field
    - **O variableLevel1**: the level 1 variable of the GCMD field
    - **O variableLevel2**: the level 2 variable of the GCMD field
    - **O variableLevel3**: the level 3 variable of the GCMD field
    - **O uuid**: the unique identifier of the GCMD field

The field **gcmdKeywords** is deprecated as it is replaced by the more generic **keywords** field.

  - **M theiaCategories**: JSON character string table. Each character string represents a URI identifying a term of type Category of variable (TYPE = Category of variable) in the Theia/OZCAR thesaurus ([https://w3id.org/ozcar-theia/c\\_341ce66a](https://w3id.org/ozcar-theia/c_341ce66a)) and enables the producer variable name to be associated with the Theia/OZCAR variable category taxonomy. It is possible to fill in several URIs if you want to associate the producer variable with several categories. Each character string must be a valid URI of a hierarchical level allowing at least to identify the compartment of the critical zone in which the measurement is carried out (e.g. for a Water level variable to be able to distinguish if it concerns karst water or groundwater). Alternatively, it is now possible to reference a variable URI rather than a category URI if this term of type "Variable" (TYPE = Variable) already exists in the thesaurus and describes the producer variable precisely enough. In this case the JSON table will contain one and only one URI representing the Theia/OZCAR variable with which to associate the producer variable.
  - **O keywords**: JSON table of JSON objects. Each of the JSON objects representing a keyword. The JSON Object representing a keyword is composed of 2 key/value pairs. The first pair with key = "keyword" is a free text field describing the keyword, and the second pair with key= "uri" is a free text field describing the URI of the keyword in a published thesaurus.- **M featureOfInterest**: is an abstraction of the real world that can be defined by spatial dimensions. This entity of interest corresponds to the target of the observation. The interest entity is represented here as a JSON object consisting of the following field:
  - **M samplingFeature**: this is the sampling entity of the measurement. This field will allow you to locate the observation in the application. This entity is described according to spatial dimensions and can be represented by a GeoJSON object composed of the following fields:

- **M name**: the name of the sampling entity. For example, the name of the measuring station for a point observation, the name of the surface for an observation concerning a surface. **It is mandatory that the chosen name be persistent in time.**
- **M geometry**: member describing the geometry of the sampling entity. The geometry object is composed of the following two members:
  - **M type**: the type of the geometry of the measurement. The possible types of influence are "Point", "MultiPoint", "LineStrings", "MultiLineStrings", "Polygon", and "MultiPolygon".
  - **M coordinates**: For a "Point" type, the coordinates member must be a position, either a longitude/latitude coordinate doublet or a longitude/latitude/elevation triplet. The coordinates are expressed in decimal degrees. For a "MultiPoint" or "LineString" type, the coordinates member is a JSON table of positions. For a "MultiLineString" or "Polygon" type, the coordinates member is a JSON table of JSON positions table. For type "MultiPolygon" the coordinates member is a JSON table of JSON table of positions. Each position is described by the coordinate doublet[longitude, latitude] or a triplet [ longitude, latitude, elevation].
- **M The constant "type": "Feature"** is assigned in order to match the object to the GeoJSON schema.
- **M The constant "properties": {}** is assigned in order to match the object to the GeoJSON schema.
- **O procedure**: observation protocol describing the treatment used to produce a result on an observed property of the object of interest. It is often provided by an instrument or sensor but can also be a processing chain, a human observer, a simulation. This object is used to document the observation acquisition protocol but also the processing carried out since the acquisition. This object is described by a JSON object composed of the following fields:
  - **O dataProduction**: JSON object used to describe the data acquisition. It is composed of the following fields:
    - **O method**: character string describing the acquisition method.
    - **R sensors**: collection of JSON objects describing the device that was used to produce the data (physical or virtual sensor). This field allows the implementation of standardized data exchange webservice (OGC SensorThings, OGC SOS). The objects in the collection correspond to one of the following 2 JSON objects:
 

**PhysicalSensor :**

      - **R model**: the sensor model
      - **R manufacturer**: the manufacturer of the sensor
      - **O serialNumber**: the serial number of the sensor
      - **M sensorType**: the type of sensor
      - **O operationalMode**: character string describing the operating mode of the sensor
      - **O calibration**: character string describing the calibration of the sensor.
      - **R activityPeriods**: collection of JSON objects with the following members:
        - **M dateBeg**: string describing a date in ISO 8601 format "YYYY-MM-DDThh:mm:ssZ".



- **M dateEnd**: string describing a date in ISO 8601 format "YYYY-MM-DDThh:mm:ssZ".
- **O documents**: collection of documents associated with the physical sensor represented by a JSON table. Documents can be of two types: "Publication" or "Manual". The **document** object is a JSON array. Each of the elements in this table is a JSON object with two members. A "type" member can take the value "Publication" or "Manual", a "url" member representing the url to which the document is accessible. This optional element is used to document reference publications associated with the physical sensor.

#### VirtualSensor :

- **M name**: the name of the virtual sensor (Ex: model name)
- **O parametrisationDescription**: the description of the forcings and parameters applied to the model
- **O documents**: collection of documents associated with the virtual sensor represented by a JSON table. Documents can be of two types: "Publication" or "Manual". The **document** object is a JSON array. Each of the elements in this table is a JSON object with two members. A "type" member can take the value "Publication" or "Manual", a "url" member representing the url to which the document is accessible. This optional element is used to document the reference publications associated with the virtual sensor.
- **O lineageInformation**: collection of JSON objects to document the post-acquisition processing of the data. Each of the items in the collection consists of the following fields:
  - **M processingDescription**: string of characters describing the processing.
  - **O processingDate**: the date on which the processing was performed. This date is written in ISO 8601 format "YYYY-MM-DDThh:mm:ssZ".
- **M result**: the result of an observation is an estimate of the value of a property of a feature-of-interest, obtained using a specific procedure. For an observation that represents time-series data, the result object represents the time series of a given variable at a given station. The **result** object is described by the JSON object composed of the following members:
  - **O missingValue**: code associated with missing values in data files.
  - **R qualityFlag**: collection of objects describing the quality flags and their meaning associated with measurements. JSON objects in the collection are composed of the following members:
    - **M code**: string describing the flag code used in the data files
    - **M description**: string of characters describing the meaning of the quality flag representing by the code.
  - **R additionalValues**: Additional values can be added to the observation measurements (examples: uncertainties, errors, parameter values related to instrumentation...). These additional values are present in the data file and are described in the pivot format by a collection of JSON objects. The JSON objects in the collection are composed of the following members:
    - **M columnName**: character string describing the name of the column of the data file where the additional values of the observation are detailed.
    - **M name**: character string describing the name of the additional value.
    - **R description**: string of characters describing the additional value
    - **O unit**: character string describing the unit of the additional value.

- **M dataFile**: object describing the data file containing the observation results. Each object is described by the following field:
  - **M name**: the name of the file with its extension. This field will allow you to find the data file in the system. The application will then be able to use the file to perform operations (for example, viewing).

**The nomenclature of file names and its format is described in the data file description document.**

- **M dataType**: the format of the observation result. This field will allow the application to perform operations on the observation such as visualization. This field can have the following values: "Numeric", "Text", "Vector", "Raster", "Photo", "Video", "Audio", "Other".
- **M timeSeries**: Boolean equal to TRUE if the observation is a series of time data. This field will allow the application to perform operations on the observation such as visualization.
- **M temporalExtent**: describes the valid time extension of the observation data. If the observation is a time series:[Start date of the series, end date of the series]. If the observation is not a time series, the validity period of the measurement according to the scientist (e. g. :
  - Measurement date, measurement date + 6 months] for a measurement that can be considered constant over 6 months.
  - Measurement date, 9999-12-31T00:00 :00Z] for a measurement that can be considered timeless. ). If the date of the measurement is unknown[9999-12-31T00:00:00:00Z, 9999-12-31T00:00:00Z]

The temporalExtent object is composed of the following members:

  - **M dateBeg** : string describing a date in ISO 8601 format "YYYY-MM-DDThh:mm:ssZ".
  - **M dateEnd**: string describing a date in ISO 8601 format "YYYY-MM-DDThh:mm:ssZ".
- **M processingLevel**: Level of processing of the observation. The object can have one of the following values: "Raw data", "Quality-controlled data", "Derived products".  
**Raw**: The data has not been subjected to quality control. **Quality-controlled data**: the data have undergone quality control such as visual or statistical inspection (e.g. correction of sensor drift, removal of outliers). **Derived products**: the data are the result of a scientific and/or technical interpretation. Data derived from scientific interpretation or based on a model that uses other data and/or is based on strong assumptions. Ex: data built from data from several sensors, gap-filled data, model data...



## Annexes 1: Observatory identifiers

Observatories	IDs
<b>AgrHyS</b>	HRMA
<b>AMMA-CATCH</b>	CATC
<b>Auradé</b>	AURA
<b>BVET</b>	BVET
<b>CRYOBS-CLIM</b>	CRYO
<b>Draix Bléone</b>	DRAI
<b>ERORUN</b>	ERUN
<b>H+</b>	HPLU
<b>HYBAM</b>	HYBA
<b>MSEC</b>	MSEC
<b>OBSERA</b>	OBSE
<b>OHGE</b>	OHGE
<b>OHM-CV</b>	OHMC
<b>OMERE</b>	OMER
<b>OPE</b>	OPEA
<b>ORACLE</b>	ORAC
<b>SRO</b>	OSRC
<b>Réal Collobrier</b>	REAL
<b>SNO Karst</b>	KARS
<b>SNO Peat bog</b>	TOUR
<b>Yzeron</b>	YZER

## Annexes 2: values of the enumerations

Possible values of the type member of an element of the **fundings** object:

<<Enumeration>> EnumFundingTypes
+ French research institutes
+ French universities and schools
+ Other universities and schools
+ Other research institutes
+ Research unit
+ French federative structure
+ Other federative structure
+ Research program
+ Other

Possible values of the member role of an element of the object **contacts**

<<Enumeration>> EnumContactPersonRoles
+ Principal investigator
+ Project leader
+ Project member
+ Data manager
+ Data collector

<<Enumeration>> EnumContactOrganisationRoles
+ Research Group

Possible values of the **dataType** object :

<<Enumeration>> EnumDataTypes
+ Numeric
+ Vector
+ Raster
+ Photo
+ Video
+ Text
+ Audio
+ Other



Possible values of the **geological** object:

<<Enumeration>> EnumGeologies
+ Volcanic rocks
+ Plutonic rocks
+ Metamorphic rocks
+ Carbonate rocks
+ Quaternary soils
+ Other sedimentary rocks

Possible values of the **climates** object:

<<Enumeration>> EnumClimates
+ Oceanic climate
+ Tropical climate
+ Mediterranean climate
+ Mountain climate
+ Continental climate
+ Polar climate
+ Arid climate
+ Equatorial climate

Possible values of the object `inspireTheme` (see: <http://inspire.ec.europa.eu/theme>)

<<Enumeration>> EnumInspireTheme	
+ Addresses	
+ Administrative units	
+ Agricultural and aquaculture facilities	
+ Area management/restriction/regulation zones and re	
+ Atmospheric conditions	
+ Bio-geographical regions	
+ Buildings	
+ Cadastral parcels	
+ Coordinate reference systems	
+ Elevation	
+ Energy resources	
+ Environmental monitoring facilities	
+ Geographical grid systems	
+ Geographical names	
+ Geology	
+ Habitats and biotopes	
+ Human health and safety	
+ Hydrography	
+ Land cover	
+ Land use	
+ Meteorological geographical features	
+ Mineral resources	
+ Natural risk zones	
+ Oceanographic geographical features	
+ Orthoimagery	
+ Population distribution - demography	
+ Production and industrial facilities	
+ Protected sites	
+ Sea regions	
+ Soil	
+ Species distribution	
+ Statistical units	
+ Transport networks	
+ Utility and governmental services	

Possible values of the **topicCategories** object (see: <http://inspire.ec.europa.eu/metadata-codelist/TopicCategory>)

<<Enumeration>> EnumTopicCategories
+ Boundaries
+ Biota
+ Climatology / Meteorology / Atmosphere
+ Economy
+ Elevation
+ Environment
+ Farming
+ Geoscientific Information
+ Health
+ Imagery / Base Maps / Earth Cover
+ Inland Waters
+ Intelligence / Military
+ Location
+ Oceans
+ Planning / Cadastre
+ Society
+ Structure
+ Transportation
+ Utilities / Communication

Possible values of the type member of an element of the **document** object:

<<Enumeration>> EnumDocumentTypes
+ Publication
+ Manual

Possible values of the **processingLevel** object :

<<Enumeration>> EnumProcessingLevels
+ Raw data
+ Quality-controlled data
+ Derived products