DMP du projet "Theia/OZCAR Information System"

Plan de gestion de données créé à l'aide de DMP OPIDoR, basé sur le modèle "ANR - DMP template (english)" fourni par Agence nationale de la recherche (ANR).

Plan Details

Plan title	DMP du projet "Theia/OZCAR Information System"
Version	Mid term version
Fields of science and technology (from OECD classification)	Earth and related environmental sciences
Language	eng
Creation date	2020-12-03
Last modification date	2022-12-12
Identifier	https://in-situ.theia-land.fr/
Identifier type	URL
License	**Name**: Creative Commons Attribution 4.0 International **URL**: http://spdx.org/licenses/CC-BY-4.0.json
Associated documents (publications,reports, patents, experimental plan...), website	• Theia/OZCAR in situ data : https://www.theia-land.fr/en/product/in-situ-data/ • Theia data pole : https://www.theia-land.fr/ • OZCAR Research Infrastructure : https://www.ozcar-ri.org/ • Braud, I., V. Chaffard, C. Coussot, S. Galle, P. Juen, and the Theia/OZCAR team (25 persons). « Building the Information System of the French Critical Zone Observatories Network: Theia/OZCAR-IS ». Hydrological Sciences Journal Special Issue: Hydrological data: opportunities and barriers (4 mai 2020): 02626667.2020.1764568. : 10.1080/02626667.2020.1764568

Project Details

Project title	Theia/OZCAR Information System
Abstract	—————————————————————————— DMP Template: Agence Nationale de la Recherche (ANR): ANR - DMP template (english) DMP Template Version: V4 released may 7 2020 Authors: S. Galle, I. Braud, V. Chaffard, C. Coussot, B. Boutherin

+ observatories data managers (see table 3)

Responsible for updating the DMP: Sylvie Galle (see section 6a)

Contact: sylvie.galle@ird.fr

Version V2020-12-11: version submitted with the mid-term report of the FairTOIS ANR project [ANR-19-DATA-0003].

Version V2022-12-15: updated version for the final report of the ANR FairTOIS project [ANR-19-DATA-0003].

Deposited on the OPIDOR portal: https://dmp.opidor.fr/public_plans?page=1&search=theia

The latest version of the DMP is visible online (readonly): https://cloud.univ-grenoble-alpes.fr/s/3drTSG3wccLDeL6

————————————————————————————

This document is the Data Management Plan (DMP) for the national inter-agency Theia/OZCAR Information System (IS) [ii] .

This IS is developed jointly by the **OZCAR** [iii] (French network of Critical Zone observatories) Research Infrastructure (RI) and the French **Theia land surface data pole** [xxv]. Theia/OZCAR is part of the **Data Terra** [v] French numerical research infrastructure “Integrated observation of the Earth system”.

This DMP applies to the continental surface in situ data, that will be made accessible, stored, shared, and redistributed within the Theia/OZCAR IS.

The aims of the Theia/OZCAR Information System (IS) (Braud *et al.*, 2020) are:

- to make all continental surface in situ data visible and easy to access,
- facilitate their discovery on a single portal,
- allow their interoperability and foster their citations

The Theia/OZCAR-IS is developed based on data exchange standards (INSPIRE, OGC) and is committed to implementing the FAIR principles [iv], especially to prepare the French community for the integration of European projects. Moreover, we follow the recommendations established in Data Citations Principles [xxiii].

Theia/OZCAR operation principles (Figure 1): A continuous information flux is organized through the implementation of the pivot data model and using vocabulary elements provided by the thesaurus. Information exchanges are provided by data producers. Since the data producers have developed different types of databases (relational database, ftp repository) and have different level of computer skills, they can either use in-house script or interoperability webservices (API) to push their data into the system. Final users (either humans or machines) can consume the services provided by the Theia/OZCAR IS for different purposes (Fig. 1).

Download figure 1 : https://cloud.univ-grenoble-alpes.fr/s/zDL6BgMK32FZJTg

Figure 1*: Scheme of the principles of the Theia/OZCAR IS: Different data producers (green rectangles) with different types of databases exchange information with the central Theia/OZCAR IS (yellow rectangle). A continuous information flux is organized through the implementation of the pivot data model and using controlled vocabulary (red arrow). Users (either humans or machines) can consume the services provided by the Theia/OZCAR IS for different purposes (blue rectangles).*

In a first step, Theia/OZCAR-IS gathers the 22 labelled long-term observatories of the OZCAR RI [iii]. The observatories document various compartments of the critical zone at approximately 75 sites, ranging from a few ha to several thousand km2 in France and abroad (Gaillardet *et al.*, 2018). Critical zone variables include physical and chemical variables. The practices and initial objectives of the observatories are different, explaining the heterogeneity in the choice of sensors, monitoring protocols and data management. In the second step, Theia/OZCAR will gather data from other sources (research programs, non-labelled observatories, official data repository) willing to make their data FAIR. The OZCAR-RI offers a representative panel of data for this second step.

The complexity of Theia/OZCAR IS does not lie in the total amount of data stored , but in the variety of data and metadata needed to contextualise them (variables, objects of interest, acquisition methods) and the heterogeneity of existing distributed information systems to describe and disseminate data from each observatory.

Observatory data are managed either by the observatories themselves or by a data centre (Appendix 1). The data centres are distributed nationally, close to the data producers.

The responsibilities of the different actors with regard to the data are as follows:

- The observatories ensure the collection of data and their scientific validation, and they determine the rules of access and the reuse of their data. It is the responsibility of the data producers to provide data that respect the GDPR law and are anonymised if required.
- The data centres ensure the documentation of the data (metadata), their medium-term storage and sometimes their dissemination via their own information system.
- The Theia/OZCAR IS ensures the homogenisation and enrichment of metadata, data visibility and access on a single portal. It ensures compliance with the diffusion rules set by producers and the traceability of data downloads. Theia/OZCAR can advise producers on the choice of an open license respecting national open data policies. Depending on the progress of Data Terra RI development, the Theia/OZCAR IS will propose to the observatories the possibility of obtaining data DOIs . Long term data archiving is not yet included in the scope.

The non-permanent staff involved in the development of the Theia/OZCAR information system are supported by CNRS-INSU, IRD, OZCAR-RI and ANR FairTOIS “Implementing FAIR principles for Theia/OZCAR In Situ data information system” (ANR-19-DATA-0003, 2020-2022)[i]. The permanent staff contributing to the development of the Theia/OZCAR IS are funded by CNRS, IRD, INRAE, BRGM, and several universities. The staff involved in the project and people’s roles are listed in section 6 (Tables 2 and 3).

Funding

- CNRS-INSU, IRD, OZCAR-RI and ANR FairTOIS : ANR-19-DATA-0003

Start date 2017-06-01

Partners

- Institut de recherche pour le développement https://ror.org/05q3vnk25
- Centre national de la recherche scientifique https://ror.org/02feahw73
- Observatoire des Sciences de l'Univers de Grenoble (OSUG)

Research outputs :

1. Theia/OZCAR Information System for continental surface in situ data

Contributors

Name	Affiliation	Roles
Boutherin Bernard	OSUG	
Braud Isabelle - https://orcid.org/0000-0001-9155-0056	INRAE	
Chaffard Véronique - https://orcid.org/0000-0003-2823-7117	IRD	• Personne contact pour les données
Coussot Charly - https://orcid.org/0000-0002-0544-4802	OSUG	
Galle Sylvie - https://orcid.org/0000-0002-3100-8510	IRD	• DMP manager • Project coordinator

DMP du projet "Theia/OZCAR Information System"

1. Data description and collection or re-use of existing data

1a. How will new data be collected or produced and/or how will existing data be re-used?

At the time of the present DMP, the Theia/OZCAR IS does not produce data but gathers existing data from several producers' information systems, making them homogeneous and conforming to international standards for open data. We explain below how the data are collected from data producers (observatories).

- Explain which methodologies or software will be used if new data are collected or produced.

A common data model, based on various metadata standards, was defined by the Theia/OZCAR technical team, in interaction with the observatories data managers to set up information fluxes between data producers' IS and the Theia/OZCAR IS. This so-called "pivot data model" was built to get all the necessary information to (i) handle various types of data; (ii) standardise data description; (iii) fill in the chosen standards metadata; and (iv) ensure the functional requirements of the system, such as the search criteria in the web interface and the implementation of standardised web services.

The pivot data model describing the mandatory, recommended or optional information is described in the Theia/OZCAR GitHub repository [vi].

Data producers must set up a tool (API or script, see fig 1) to (i) extract and format the information from their existing IS and (ii) upload it to the Theia/OZCAR-IS dedicated storage. At the time of the present DMP, only the script is operational. An API using interoperability webservices (OGC Catalog Service for the Web and OGC SensorThings API) is being defined with the interested data producers

The tools must generate an archive, compliant with the pivot data model, that contains all the metadata associated with time series measurements (see detail section 1.b). Data and metadata are uploaded to Theia/OZCAR IS using an HTTP PUT request method. Once validated, the content of the JSON files is imported into a MongoDB database, from which it is processed to feed the user interface and to provide web services. The data files are stored in the file system.

This data flux is organized as a continuous workflow. The frequency of updating the JSON files is left to the data producers. At the time of the present DMP, when a new version of a JSON file is transmitted to the Theia/OZCAR IS, it must include all the data of the data producer, and previous information is deleted and replaced by the new one in the Theia/OZCAR IS MongoDB database. Therefore, a copy of the information available in the Theia/OZCAR IS is kept up to date and can be considered as a data cache.

At the time of the present DMP, the version management is not specifically taken into account, but in the future a software component has to be developed on this important subject in coordination with Data Terra RI.

To ensure that the information available in the Theia/OZCAR IS is always up to date, data producers are advised to automate the transmission of the JSON + data files.

- State any constraints on reuse of existing data if there are any.

Metadata are in open access. No restrictions apply to metadata, even for data acquired by some observatories that are not open (for example, data on pesticides, farmers' practices, etc.)

Data: Theia/OZCAR respects the data policy or licence of the data producers. The Theia/OZCAR portal only allows downloading of data in full open access or with the following limitations: (i) embargo period and/or (ii) credit/acknowledgments requests (e.g., CC-BY). For other data (sensitive data), data producers must be asked to get the data.

- Explain how data provenance will be documented.

The data provenance is provided in the pivot data model. The mandatory fields include the name of the observatory to which the dataset is attached, its funding institution(s) (name and type), the scientist responsible for the dataset, the head of the observatory (name and research unit) and the DOI (if available).

- Briefly state the reasons if the reuse of any existing data sources has been considered but discarded

In Theia/OZCAR IS, no data source is discarded provided it belongs to the continental surface scientific domains.

1b. What data (for example the kind, formats, and volumes), will be collected or produced?

Data are of different kinds: time series, vectorial and raster maps, and 2D profiles. The total volume is not large, but it includes a large number of very different data sets.

- Give details on the kind of data: for example numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media.

Most of the produced data are time series observations acquired by sensors in the field, generally at one location. However, some observations are collected on more complex geometries, such as curves or surfaces (e.g., glacier mass balance is computed using drifting sensors providing data on a trajectory).

Other types of data are also identified. This includes vectorial and raster maps (land use, land cover, digital elevation model, and soil maps), ex situ data with observations made on samples (e.g., soil cores, soil or water samples), and 2D profiles describing a deep soil structure using different geophysical techniques. Some observatories produce spatialized products derived from a combination of several data sources, such as interpolated rainfall fields from rain gauges (Vischel *et al*. 2011) and reanalyses combining radar rainfall and rain gauges (Boudevillain *et al*. 2016), remote sensing products such as soil moisture fields (Pellarin *et al*. 2009), and model results. Additional types of data include datasets describing agricultural practices, such as the application of fertilizers or pesticides, surveys, interviews, photos, or videos.

This is the potential. Not all of these data types are implemented at the time of the present DMP (see section 1.a). At the time of the present DMP only time series data are handled. A future objective is to handle all types of data.. Rather than undertake all the development, Theia/OZCAR SI has chosen to rely on community projects being developed and to harvest them. For example, a specific data base dedicated to hydrogeophysics data is developed at Geosciences Rennes (https://hplus.ore.fr/en/database/acces-database) which stores this kind of data for all the OZCAR-RI observatories.

- Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf).

Each data producer produces a JSON file compliant with the pivot data model, which contains all the metadata associated with the time series measurements. Time series measurements must be provided as CSV files (one per variable at a given location) and packaged into .zip archives (one per dataset). The format of the CSV file was standardised with a common syntax: the data file must have a short metadata header, and data values must be formatted in the form of a tuple, with date, location, value, and quality flags. Data files are stored on a file system, and metadata are stored in a MongoDB database.

In the future, JSON file compliant with pivot data model may also be derived from information provided by interoperability webservices. In this case, Metadata documented in the JSON file will be stored in a MongoDB Database, whereas webservice for data dissemination (OGC SensorThings API) provided by data producer IS will be used to provide services on data in Theia.OZCAR IS.

- Justify the use of certain formats. For example, decisions may be based on staff expertise within the host organisation, a preference for open formats, standards accepted by data repositories, widespread usage within the research community, or on the software or equipment that will be used.

Theia/OZCAR project favours open and easy-to-use formats and softwares:

For collecting the data from the observatories, we chose

- the CSV format (with a normalised header) for the time series data files
- a JSON format for the metadata file (with a GeoJSON format for geographical features).

The CSV format was chosen because we handle a small amount of data, as it is a self-documented format and is widely used by different communities. The choice of JSON implementation was motivated by the following reasons: JSON is easily read and understood by humans (as opposed to XML), JSON is less verbose than XML, and like XML, JSON files can be validated using schemas (JSON Schema, IETF 2019).

The MongoDB database was chosen to store the metadata, as it is an open-source, document-oriented, schema-less database that facilitates its evolution if new data types have to be included. It is adapted to store, retrieve, and manage JSON semi-structured data.

Downloading the time series from the web interface in a common, standardised, self-documented, open, and easy-to-use format is planned, but has not been developed at the time of the present DMP. Both CSV (with normalised header) and NetCDF (compliant with the CF convention) formats are targeted.

- Give preference to open and standard formats as they facilitate sharing and long-term reuse of data (several repositories provide lists of such ‘preferred formats’).

The main criteria used for choosing formats: open, widely used, well-defined (e.g., based on ISO standards) and preferably self-documented.

- Give details on the volumes (they can be expressed in storage space required (bytes), and/or in numbers of objects, files, rows, and columns).

The volume of the time series data that will be collected by the Theia/OZCAR project is moderate (approximately 10 TB). This could increase in the future (beyond the ANR FairTOIS project) with the development of high frequency/high resolution sensors and the inclusion of model simulation data.

More than 10 000 different observations are referenced, described, and stored in the system at the time of the present DMP. An observation represents a time series measurement of a property made repeatedly with a defined procedure in a location (or more generally on a “feature of interest"); see Figure 2.

---

2. Documentation and data quality

2a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany the data?

- Indicate which metadata will be provided to help others identify and discover the data.

The pivot data model was defined to describe the metadata. It includes four main concepts: data producer, dataset, observation and result (see Figure 2 below). The data are organized in datasets.

- A data producer or observatory is defined by its Id, name, title, description, funding institutions and contacts.
- A dataset is a collection of observations. The information describing datasets (what, where, when, who, how, data use and access condition) was based on the ISO 19115/INSPIRE standard [vii] and the DataCite standard [viii]. These metadata are required to identify the dataset and its “institutional context”.
- An observation (formally a collection of observations) is a serie of measurements, at a specific time, of a property value (observed variable). This property value is referring to a feature-of-interest represented by its sampling feature (station or a specimen) and obtained using a specified procedure (physic/virtual sensors or observers). The information describing the observation data aims to describe what it relates to and how it was acquired to allow its reuse. They are based on the O&M OGC standard [ix] (OGC, 2013). They consist of variable names and units (ObservedProperty concept), features of interest involved in observations (FeatureOfInterest concept), acquisition and processing methods (Procedure concept), and observation values (Result concept).
- The Result concept refers to the data itself. It contains some metadata on the data values (quality flag description, missing value and the data filename. For an observation that represents time-series measurements (of a property made repeatedly with the same procedure in a feature of interest), the result comprises a single .csv file, which contains only one time series of a given variable at a given location.

See figure 2 online : https://cloud.univ-grenoble-alpes.fr/s/exzDpQSQozM7K78

Figure 2*: Simplified visualization of the pivot data model. Information in the blue rectangle corresponds to the ISO19115/Inspire information, while that in the red rectangle corresponds to the O&M (Observations and Measurements) standard (after Braud et al., 2020).*

- Indicate how a common variable vocabulary was chosen to help others identify and discover the data.

Theia/OZCAR has defined its own vocabulary for variable names. It uses existing vocabularies or standards for the following information: funding structures, names of contributors, and various other information (date, contact details, etc.). These choices are detailed below.

Variable names

Harmonisation of the vocabulary was necessary to make the data presented by the Theia/OZCAR IS easily discoverable. Thus, a FAIR thesaurus of variable names has been formalised to aggregate and harmonise the data from the different observatories entering our system.

A hierarchised controlled vocabulary was constructed based on categories and variable names of the Global Change Master Directory (GCMD[x]). Launched by NASA in 1987, the GCMD covers subject areas within the Earth and environmental sciences. Today, it is one of the largest public keyword inventories in the world. Its use is recommended in the French Data Terra IR. As it does not include chemical data, it was enriched using the SANDRE [xi] vocabulary for chemical data (the French norm for data reporting for the Water Framework Directive [xxvi]). Additional categories relevant for the OZCAR-RI community, such as “surface fluxes” has also been added.

The Theia/OZCAR controlled vocabulary [xvi] consists of hierarchised concepts of variable categories and names, covering the different compartments of the critical zone, where variable names are found at the last level of the hierarchy. A given variable or a category may be related to several compartments (e.g., “surface flux” was related to both “land surface” and “atmosphere”).

The variable names used for the search are generalists, but, once found, it can be useful for the user to know additional information (depth of measurement, time step, etc.). The I-ADOPT framework is chosen to describe and harmonise all available information about a variable. This framework is recommended by the Research Data Alliance.

I-ADOPT provides a simple ontology enabling variable name modelling according to four classes. Complex variable name are split into atomic elements (Figure 3) which foster semantic alignments with other vocabularies (Coussot et al. 2022). Statistical vocabulary elements used in variable name are described using Complex Property Model (CPM) ontology [xxxi], following the INSPIRE Guidelines D2.9 [xxx]. Each complex variable name is linked to a simplified label which is used for the discovery of the data on the web portal.

See figure 3 online: https://cloud.univ-grenoble-alpes.fr/s/FdTZn2TTrw2MySK

Figure *3: Decomposing a complex variable in the I-Adopt framework: case of “One day cumulative liquid precipitation amount”. The variable is composed of (i) an ObjectOfInterest, the precipitation, (ii) a Property, the volume, and completed by (iii) a Constraint applied on the ObjectOfInterest, liquid, (iv) a statistical aggregation and (v) the time period on which is performed the statistical aggregation, 1 day.*

The data producers keep their own variable names. Each producer name is then linked to the common controlled vocabulary. The pairing was proposed by the scientists of the project team, and data producers were then consulted to validate the choices.

At the time of the present DMP, only the variables actually in the information system t are listed on the Theia/OZCAR web interface. Their number is approximately 300, with approximately 45% related to chemical data.

Funding structures

Their description and type are identified and harmonised using the scanR database [xii]. scanR is a service offered by the French Ministry of Higher Education, Research and Innovation. It allows searching for companies that are active in research and innovation, public research laboratories, public funding, research work.

Role of the contributors

Contributor types from the DataCite standard [viii] and the ISO 19115 standard Responsible Party Role (principal Investigator) [xxvii] are used. As far as it exists, the contributors are identified by their ORCID number.

Other information

- Spatial coordinates: spatial reference system WGS84 (EPSG 4326);
- Time: ISO 8601 standard [xxviii];
- Topics and themes: ISO 19115 topic categories [xiii] and INSPIRE spatial data themes [xiv] are used in the pivot data model to tag datasets;
- Country: code ISO 3166 standard [xxix] is used in the pivot data model to identified the country of the funding or contributor organisation;
- Geology: unit legend of the World CGMW 1:50M Geological Units Onshore map;
- Climate: climate classes of the Köppen climate classification.

- Indicate the link between Theia/OZCAR vocabulary with international semantic thesauri.

The Theia/OZCAR thesaurus is published according to Semantic Web standard. It is published on the web using the SKOS (Simple Knowledge Organisation System) Standard and published on the Web using Skosmos (Suominen et *al.*, 2015) [xv] . It can be browsed and downloaded following this link: https://w3id.org/ozcar-theia [xvi].

To promote semantic interoperability of the data, mapping was performed with international semantic thesauri. It links the terms of the Theia/OZCAR thesaurus with exact or similar terms of international thematic thesauri (see Table 1 below for details on the thesauri considered). This mapping was performed for each variable name and category name.

Table 1: Thesauri that were used to map the Theia/OZCAR category names on existing ontologies. The same thesauri will be used to map the variable names (after Braud et al., 2020)

Ontology names	Developed by	Field of interest	Web link
GCMD (Global Change Master Directory)	NASA (USA)	Earth Science	https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords
GEMET (GEneral Multilingual Environmental Thesaurus)	EEA (European Environment Agency)	Environment	https://www.eionet.europa.eu/gemet/en/themes/
EnvThes (Environmental Thesaurus)	LTER-Europe (EU)	Environment	http://www.enveurope.eu/news/envthes-environmental-thesaurus
AGROVOC*	FAO (UN)	Agriculture (FAO)	http://aims.fao.org/fr/agrovoc
AnaEE* (ANAlysis And Experimentation Ecosystems)	ANAEE-RI (EU)	Continental ecosystems and their biodiversity	https://lovinra.inra.fr/2017/03/13/thesaurus-anaee/
ChEBI* (Chemical Entities of Biological Interest)	ELEXIR infrastructure (EU)	Chemistry	https://www.ebi.ac.uk/chebi/downloadsForward.do
NetCDF* (Climate and Forecast Standard Names)	Climate and Forecast Standard Names Committee	Earth science	http://vocab.nerc.ac.uk/collection/P07/current/

** : these ontologies are not aligned at the time of the present DMP*

- Indicate which metadata standards (for example DDI, TEI, EML, MARC, CMDI) will be used.

We considered the following standards:

- **ISO19115/INSPIRE**, as it is necessary to set up standardised metadata exchange services.
- **O&M** (Observations and Measurements), as it is necessary to set up standardised data exchange services.
- **Datacite**, as it is used by DOI registration services.
- **schema.org** [xviii], as it allows datasets to be referenced by Google Dataset Search.
- **DCAT** [xix], as it allows facilitating interoperability between data catalogs published on the Web.
- **GeoDCAT** is being considered for building Data Terra catalog.
- **SensorThings** API is based on O&M and is set up for data ingestion and dissemination.

To build the Theia/OZCAR pivot data model, mapping between these standards was performed. The mapping between the various standards is provided in a dedicated data model documentation directory of the project's GitHub repository [xx]. A complete description can be found in the directory of the data model documentation in the project GitHub repository [xxi].

- Use community metadata standards where these are in place.

Table 1 above lists the Thesauri that were mapped to the Theia/OZCAR thesaurus

- Indicate how the data will be organised during the project, mentioning for example conventions, version control, and folder structures. Consistent, well-ordered research data will be easier to find, understand, and reuse.

Data files are stored on a file system organised by producers, and metadata are stored in a MongoDB database.

- Consider what other documentation is needed to enable reuse. This may include information on the methodology used to collect the data, analytical and procedural information, definitions of variables, units of measurement, and so on.

This information is included in the pivot data model. The location (feature of interest) and sampling feature, the name of the variable, and the unit are mandatory fields. The protocol is an optional field at the time of the present DMP.

- Consider how this information will be captured and where it will be recorded for example in a database with links to each item, a 'readme' text file, file headers, code books, or lab notebooks.

When downloading the data, all the metadata necessary to contextualise the data will be supplied on the data file headers. At the time of the present DMP, the definition of the content is under discussion with the scientific users.

2b. What data quality control measures will be used?

- Explain how the consistency and quality of data collection will be controlled and documented. This may include processes such as calibration, repeated samples or measurements, standardised data capture, data entry validation, peer review of data, or representation with controlled vocabularies.

Scientific data quality control is performed by data producers. The pivot data model contains metadata describing the data quality. This is done using an enumeration list (raw data, quality-controlled data, and derived product) and a data quality flag defined by data producers (code and description). These fields are optional at the time of the present DMP and could become mandatory in the future. The definition of the quality code currently depends on the observatories and should be harmonised in a second step. For time series, data producers also have the option of providing confidence intervals for their data in the data files (e.g., columns providing min-max uncertainty values) and describe their calibration procedure.

On the Theia/OZCAR IS side, each data deposit triggers a process that validates the transmitted information and stores the data. The format of the JSON file and the data files are checked for possible errors (using the JSON schema and regular expressions for .csv files). The data producers are informed in real time if the validation and import of the data into the Theia/OZCAR IS is successfull or not.

In the future, validation error details will be automatically communicated to the data producers so they can correct their files. We aim to design and run automatic verifications of the metadata to find and report 1) missing, 2) incomplete, 3)

inconsistent and 4) unrealistic metadata. Detection criteria will be improved over time.

---

3. Storage and backup during the research process

3a. How will data and metadata be stored and backed up during the research?

Theia/OZCAR's data storage and backup is based on OSUG-DC and the SUMMER Grenoble Alpes University (UGA) site infrastructure, which has been designed to offer high availability services.

At the time of the present DMP, the information system is under development and not all data producers have developed their interfacing tool. Thus, we do not yet well know the needs of Theia/OZCAR IS in terms of the volume, data security and performances, which is why we have allocated space for the project on the three storage systems described below. At the time of the present DMP, only the SUMMER system is used.

- Describe where the data will be stored and backed up during research activities and how often the backup will be performed. It is recommended to store data in least at two separate locations.

Three types of storage solutions available at OSUG-DC can be used for Theia/OZCAR data, depending on the needs of the project. The types of storage encompass notions of performance, data security and data volume. The first two solutions largely cover a high-level data security expectation, and the third allows us to go further in volume (in the short term because the space is already available).

The different means implemented to secure the data are detailed below. These means include the level of redundancy of the data on the disks, the redundancy on different geographically distant sites, the number and lifetime of the snapshots available, and the existence of a backup of the data.

1. SUMMER NFS storage is provided by UGA University, and is a very strong architecture based on Netapp@ servers. SUMMER provides redundancy of the data on two geographically distant sites. At the time of the present DMP, SUMMER is the unique storage used for the project, and 10 Tb are available on SUMMER for the Theia/OZCAR IS, which can be extended widely. At the time of the present DMP, SUMMER S3 storage must be tested and considered to store Theia/OZCAR IS data in a cloud-native way.
2. VSAN storage is provided by OSUG-DC. It is designed for the highest performance. 10Tb are available for the Theia/OZCAR project at the time of the present DMP. VSAN is distributed on 5 servers located in two geographically distant sites. VSAN redundancy is 1 for 1, which supports the cut-off of one site. VSAN does not support snapshot. Data are backed up on a large volume storage (see below).
3. A large volume storage of 64 Tb provided by OSUG-DC is available for the project. This volume offers RAID6 redundancy and is located on a single site.

- Give preference to the use of robust, managed storage with automatic backup, such as provided by IT support services of the home institution. Storing data on laptops, stand-alone hard drives, or external storage devices such as USB sticks is not recommended.

Backups are programmed in an automatic way on SUMMER storage. OSUG-DC and SUMMER's IT supports are in charge of the maintenance of the servers.

At the time of the present DMP, SUMMER is configured to have hourly, daily, weekly, monthly snapshots up to 3 months; moreover, in the event of a crash of the server itself, we have a daily backup on another server located in a different data centre.

The above backups are for data, and VMs are also backed up daily.

3b. How will data security and protection of sensitive data be taken care during the research

- Explain how the data will be recovered in the event of an incident.

In the event of an incident, data will be restored from the snapshot.

In case a backup server located in another DC is needed, the operation of restoration will be driven by the site's IT team. Note that for the 8 years that the SUMMER solution has existed, there has never been a need to use these backups.

- Explain who will have access to the data during the research and how access to data is controlled, especially in collaborative partnerships.

At the time of the present DMP, data are not available in the Theia/OZCAR data portal. In the future, the data will be available through authentication (see section 5a).

- Consider data protection, particularly if your data is sensitive for example containing personal data, politically sensitive information, or trade secrets. Describe the main risks and how these will be managed. Explain which institutional data protection policies are in place.

From a hardware point of view, the SUMMER system offers all the necessary authentication protocols to develop an authentication service, as mentioned in section 5a.

At the moment, there are no sensitive data in our DC and no need for data encryption, but if necessary, several mechanisms are available as described below:

The data stored on the SUMMER platform can be protected by several mechanisms that prevent any unauthorised modification or disclosure of data in NAS and SAN environments:

- The data centres that host the data are all equipped with access control.
- Each storage node is protected by filtering equipment (i.e., firewall), which follows a precise and adapted security policy.
- Each storage volume, whether in SAN or NAS mode, is protected by the security mechanisms linked to the ad hoc protocol (export policy, target iSCSI,) and by authentication (LDAP, Kerberos).
- The solution offers full or partial data encryption without compromising storage efficiency or performance. It is also compatible with the GemaIto SafeNet KeySecuret encryption key device, which simplifies and strengthens key management in the long term. It complies with FISMA, HIPAA, PCI, Basel li, SB 1386 and European directive 95/46/CE on data protection. At the time of the present DMP, data encryption is not activated on our volumes.

The SUMMER platform can therefore ensure data integrity, confidentiality, and authenticity.

The SUMMER platform is operated by a technical committee made up of 10 members, including an outside operator. The committee is in charge of keeping the platform in operational conditions at all times.

From a software point of view, data with restricted access come with the list of users that have access to it. Privileged users that can access particular restricted data will be identified using credentials of the authentication mechanisms that are not defined at the present time of the DMP.

---

4. Legal and ethical requirements, code of conduct

4a. If personal data are processed, how will compliance with legislation on personal data and on security be ensured?

Recommandations:

Ensure that when dealing with personal data protection laws (for example GDPR) are complied with:

- Gain informed consent for preservation and/or sharing of personal data.
- Consider anonymisation of personal data for preservation and/or sharing (truly anonymous data are no longer considered personal data).
- Consider pseudonymisation of personal data (the main difference with anonymisation is that pseudonymisation is reversible).
- Consider encryption which is seen as a special case of pseudonymisation (the encryption key must be stored separately from the data, for instance by a trusted third party).
- Explain whether there is a managed access procedure in place for authorised users of personal data.

The Theia/OZCAR IS only stores data that are provided and quality checked by data producers. It is the responsibility of data producers to provide data that respect the GDPR law.. Data producers commit to providing anonymised data where appropriate to ensure compliance with the laws and rules of confidentiality.

A specific field in the pivot format will be added to declare whether the dataset is confidential, and its access must be restricted (according to ISO 19115). When dealing with personal data, the data producers will be asked to complete a questionnaire to certify by themselves that the legal provisions are well met on various points, such as consent, the anonymization method, and the agreement of an ethics committee.

There is a plan to set up an authentication system for accessing data downloaded from the Theia/OZCAR IS (see section b). To comply with the GDPR principles, (i) the authentication information from users will be stored and protected, and (ii) users will be allowed to access and rectify their personal information and suppress their accounts.

The Theia/OZCAR web portal also presents personal information concerning the PIs of the data sets. We plan to implement a right to information, but the other rights (suppression) pose problems regarding the lineage of the data sets over time. They will be addressed later, in coordination with Data Terra policy and in compliance with the guidelines of research institutes and universities [xxiv].

4b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?

Recommandations:

- Explain who will be the owner of the data, meaning who will have the rights to control access:
- Explain what access conditions will apply to the data? Will the data be openly accessible, or will there be access restrictions? In the latter case, which? Consider the use of data access and reuse licenses.
- Make sure to cover these matters of rights to control access to data for multi-partner projects and multiple data owners, in the consortium agreement.
- Indicate whether intellectual property rights (for example Database Directive, sui generis rights) are affected. If so, explain which and how will they be dealt with.
- Indicate whether there are any restrictions on the reuse of third-party data.

The data provided through the Theia/OZCAR IS remain the property of the data producers. Data producers are also responsible for defining the embargos or restriction of use of their data. The attribution of a licence to the dataset will be encouraged by the Theia/OZCAR management team by educating the data producers. As a long-term objective, all datasets diffused by Theia/OZCAR IS will have a licence.

The data producer must be credited for the original creation and Theia/OZCAR for the diffusion of the data. This is ensured by a mandatory field of the pivot data model, where data producers must provide information about data use and constraints, and where an acknowledgment sentence is provided.

In the present version of this DMP, we do not consider the production of new data from the data stored in the Theia/OZCAR IS.

4c. What ethical issues and codes of conduct are there, and how will they be taken into account?

Recommandations:

- Consider whether ethical issues can affect how data are stored and transferred, who can see or use them, and how long they are kept. Demonstrate awareness of these aspects and respective planning.
- Follow the national and international codes of conducts and institutional ethical guidelines, and check if ethical review (for example by an ethics committee) is required for data collection in the research project.

In the unlikely situation where an ethical concern would arise from publicly sharing a given dataset (e.g., data sharing could lead to harm to people or species), the data producer will be responsible for consulting an ethics committee and/or requesting that the data access is restricted. Such restrictions will/may be set up as a temporary protective measure, pending decisions.

5. Data sharing and long-term preservation

5a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?

The objective of Theia/OZCAR is to provide FAIR data of continental surfaces to increase its sharing and reuse. The data provided will comply with data producer policies, including producer embargoes (if any). The embargo periods generally correspond to the time necessary for producers to publish their data (1-3 years).

- Explain how the data will be discoverable and shared (for example by deposit in a trustworthy data repository, indexed in a catalog, use of a secure data service, direct handling of data requests, or use of another mechanism).

The data will be discoverable and downloadable on the Theia/OZCAR web portal [xxii]: https://in-situ.theia-land.fr/

The data will also be indexed in a Geonetwork metadata catalog. This allows the catalog of data to be harvested automatically through the CSW webservice.

The data download service will provide users with transparent access to data in identical export formats, regardless of data origin. At the date of the present DMP, only a link to the data producer’s web portal is provided.

The user authentication will be performed using Data Terra Single Sign On authentication to ensure that embargo and restriction of access to some data are respected. By the time of the present DMP, the solution has been successfully tested but is not yet operational.

- Outline the plan for data preservation and give information on how long the data will be retained.

The data handled within the framework of the Theia/OZCAR project are, in principle, already stored and backed up in the data infrastructure of the data producers.

In any case, as mentioned in paragraph 3, Theia/OZCAR will store and back up the data in the OSUG/UGA infrastructure.

Storage funding and associated costs are assured for 7 years (2020-2026) if the data volume remains within the expected bound at the time of the present DMP. Later, or if the data volume explodes, additional funding will have to be found. Since the infrastructure is included in an OSU and the project belongs to a national initiative, it should be possible to obtain this funding (see section 6b).

- Explain when the data will be made available. Indicate the expected timely release. Explain whether exclusive use of the data will be claimed and if so, why and for how long. Indicate whether data sharing will be postponed or restricted for example to publish, protect intellectual property, or seek patents.

Data are generally already available from the web portals of data producers.At the present time of the DMP, data download though Theia/OZCAR portal is planned but not yet available.

- Indicate who will be able to use the data. If it is necessary to restrict access to certain communities or to apply a data sharing agreement, explain how and why. Explain what action will be taken to overcome or to minimise restrictions.

The metadata provided by the data producers describe information on data licences and acknowledgments, as well as the wishes of the data producers in terms of an embargo or data diffusion.
User authentication will be implemented to manage access rights to datasets with an embargo, or to manage access granted only to some identified users. Authentication will also allow access for authenticated data producers to statistics about data downloading, which is a criterion required by funding institutions.

5b. How will data for preservation be selected, and where data will be preserved long-term (for example a data

repository or archive)?

Recommandations:

- *Indicate what data must be retained or destroyed for contractual, legal, or regulatory purposes.*
- *Indicate how it will be decided what data to keep. Describe the data to be preserved long-term.*
- *Explain the foreseeable research uses (and/or users) for the data.*
- *Indicate where the data will be deposited. If no established repository is proposed, demonstrate in the data management plan that the data can be curated effectively beyond the lifetime of the grant. It is recommended to demonstrate that the repositories policies and procedures (including any metadata standards, and costs involved) have been checked.*

In the long-term, given the wealth of these long-term environmental data, it will be necessary to question the need for very long-term archiving, which is not currently supported by data producers. Long-term preservation (for use within 50 to 100 years) will be handled in the framework of the Data Terra Research Infrastructure.

Not all the data will be long term preserved. The Theia/OZCAR project will consider the environmental cost of this long-term storage versus the potential scientific interest of the data.

Long-term data preservation is considered a continuous process. At the technical level, the storage of the data in a site infrastructure guarantees the preservation of these data over time (cf: $3).

5c. What methods or software tools are needed to access and use data?

Recommandations:

- *Indicate whether potential users need specific tools to access and (re-)use the data. Consider the sustainability of software needed for accessing the data.*
- *Indicate whether data will be shared via a repository, requests handled directly, or whether another mechanism will be used?*

To access the data from the web portal, a standard up-to-date web browser will be necessary. The format of the data is open (CSV, NetCDF) and does not require any commercial software.

5d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

At the time of the present DMP, some datasets already have a DOI. The DOI is one of the metadata fields included in the pivot data model.

Assigning DOI, through the workflow of data imported into Theia/OZCAR IS, will be considered in the long term. This kind of service is planned and will be implemented by using the DataCite API through Data Terra services when available. .

The special feature of Theia/OZCAR datasets is that they are dynamic, i.e., there is a continuous recording of a series of variables at a given location. At the time of the present DMP, the Theia/OZCAR policy on DOI is envisaged as follows:

- Have a DOI declared to Datacite for each dataset.
- Declare DOIs with rich metadata (including all information on data producers, data funding, data quality, etc.).
- A single DOI for a dynamic dataset but with the possibility of using fragments for indicating the variants of a dataset (i.e., temporal subperiod, version of the dataset).
- The DOI on the datasets can be declared either by the producers (its traceability is ensured in the pivot format) or by Data Terra (option not implemented at the time of the present DMP).

- Explain how the data might be reused in other contexts. Persistent identifiers should be applied so that data can be reliably and efficiently located and referred to. Persistent identifiers also help to track citations and reuse.

The objective is to allow the best possible description of the datasets and allow reuse by a "naive" user from another discipline. One point of vigilance is the choice of the “best” granularity of the datasets: the datasets must be fine enough to allow for a detailed description of the data, but sufficiently encompassing to associate interdependent variables. It may also be necessary to declare a DOI for the observatories and the observation sites.

- Indicate whether a persistent identifier for the data will be pursued. Typically, a trustworthy, long-term repository will provide a persistent identifier.

Theia/OZCAR will use the DOI services that the Data Terra RI plans to set up.

---

6. Data management responsibilities and resources

6a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?

- Outline the roles and responsibilities for data management/stewardship activities for example data capture, metadata production, data quality, storage and backup, data archiving, and data sharing.Name responsible individual(s) where possible.

Data production and its scientific quality are ensured by the scientists of the observatories (data producers). The data management tasks are ensured by IT teams at both levels, i.e., the observatory and Theia/OZCAR level.

- In each observatory, data managers (Table 3) are responsible for metadata production, data curation, local storage, backing up and sharing of the data. The metadata production includes preparing data in accordance with the Theia/OZCAR pivot format and depositing it on the Theia/OZCAR-IS dedicated data repository via the HTTP protocol.
- The Theia/OZCAR IT team (Table 2) is responsible for data sharing and its interoperability with data infrastructures at the national or European level. Theia/OZCAR procedures ensure the conformity of the data transmitted.

- For collaborative projects, explain the co-ordination of data management responsibilities across partners.

Two thematic researchers (Sylvie Galle and Isabelle Braud) and a technical manager (Véronique Chaffard) coordinated the Theia/OZCAR project. Data management is the responsibility of the technical manager.

The project team is in constant contact with the correspondents of the observatories (data managers). At the time of the present DMP, 22 OZCAR-labelled observatories are committed to providing their data (see Table 3). In the future, new projects may be added.

Table 2a: Theia/OZCAR and FairTOIS project team - Active team

Name	Forename	Employer/UMR, town	Function(s)
Galle	Sylvie	IRD/ IGE, Grenoble	Scientific co-manager of Theia/OZCAR; member of the Data Terra Science WG; ANR FairTOIS manager
Braud	Isabelle	INRAE/ RiverLY, Lyon	Scientific co-manager of Theia/OZCAR; co-manager of OZCAR RI; ANR FairTOIS partner
Chaffard	Véronique	IRD/ IGE/ OSUG, Grenoble	Technical manager of Theia/OZCAR; member of the Data Terra Technical WG; ANR FairTOIS technical manager
Boutherin	Bernard	CNRS/ OSUG, Grenoble	Technical Manager of OSUG-Data Centre; Theia/OZCAR and ANR FairTOIS partner
Cailletaud	Rémi	CNRS/ OSUG, Grenoble	Technical team of OSUG Data centre: network and administration system; ANR FairTOIS
Coussot	Charly	IRD/ OSUG, Grenoble	Technical team of Theia/OZCAR; design, development and operational monitoring ; Recruited to a permanent position from 01/01/2023

Table 2ab Theia/OZCAR and FairTOIS project team - Former collaborators

Name	Forename	Employer/UMR, town	Function(s) and years
Bourdaret	Jean-Marc	fixed term contract, IRD/IGE Grenoble	Technical team of the ANR FairTois - interoperability (13/10/2020-31/04/2021)
Tessane	Adrien	fixed term contract, INRAE/Riverly Lyon	Technical team of the ANR FairTois - push BDOH to Theia/OZCAR (01/11/2020-31/8/2021)
Coussot	Charly	fixed term contract, IRD, CNRS and ANR/ IGE and OSUG, Grenoble	Technical team of Theia/OZCAR - development of the information system and its functionalities (1/10/2017-31/12/2022; including ANR FairTOIS (11 months)

Table 3: Theia/OZCAR observatories data managers - Active team

Name	Forename	Employer/UMR, town	Observatory data base
Adam	Mario	INRAE/ SAS, Rennes	OZCAR-AGRHYS
Chaffard	Véronique	IRD/ IGE/ OSUG, Grenoble	OZCAR-AMMA-CATCH-DB
Branger	Flora	INRAE/ Riverly, Lyon	OZCAR-BDOH: scientific director; ANR FairTOIS, OTHU/ YZERON
Probst	Anne	CNRS/ Ecolab, Toulouse	OZCAR-Auradé
Perreal	Guillaume	INRAE/ DSI, Lyon	OZCAR-BDOH: technical director
Raidelet	Nicolas	INRAE/ DSI, Lyon	OZCAR-BDOH: technical director
Le Bouteiller	Caroline	INRAE/ OSUG, Grenoble	OZCAR-BDOH: Draix-Bléone
Legoût	Cédric	UGA/ OSUG, Grenoble	OZCAR-BDOH: Draix-Bléone
Blanchouin	Arnaud	INRAE/ HYCAR, Antony	OZCAR-BDOH: ORACLE
Delaigue	Olivier	INRAE/ HYCAR, Antony	OZCAR-BDOH: ORACLE
Tolsa	Mathieu	INRAE/ RECOVER, Aix en P.	OZCAR-BDOH: Réal Collobrier
Gabillard	Santiago	BRGM/DSI, Orléans	OZCAR-BRGM Ades
Grellet	Sylvain	BRGM/ DSI, Orléans	OZCAR-BRGM Ades

Six	Delphine	CNAP UGA/ IGE, Grenoble	OZCAR-Cryobsclim
Cordier	Emmanuel	Univ. La Réunion/ OSU Réunion, La Réunion	OZCAR-ERORUN
Payen	Guillaume	Univ. La Réunion/ OSU Réunion, La Réunion	OZCAR-ERORUN
Kbidi	Victor	Univ. La Réunion/ OSU Réunion, La Réunion	OZCAR-ERORUN
Battais	Annick	CNRS/ Geoscience, Rennes	OZCAR-H+
Kenshilikova	Meruyert	CDD / Geosciences, Rennes	OZCAR-H+
Martinez	Jean-Michel	IRD/ GET, Toulouse	OZCAR-HYBAM
Fabre	Juliette	CNRS/ OSU OREME, Montpellier	OZCAR-KARST
Lobry	Olivier	CNRS/ OSU OREME, Montpellier	OZCAR-KARST
Saurel	Jean-Marie	CNRS / IPGP, Paris	OZCAR-Obsera
Dassin	Lucas	CDD/ IPGP, Paris	OZCAR-Obsera
Herrmann	Agnes	/ Lhyges, Strasbourg	OZCAR-OHGE
Boudevillain	Brice	CNAP/ IGE, Grenoble	OZCAR-OHM-CV
Ayral	Pierre-Alain	/ UMR ESPACE, Alès	OZCAR-OHM-CV
Nord	Guillaume	CNAP/ OSUG, Grenoble	OZCAR-OHM-CV
Dubreuil	Arnaud	INRAE/ LISAH Montpellier	OZCAR-OMERE
Baillion	Philippe	CESBIO Toulouse	OZCAR-OSR-SW
Bissi-Yandia	Thierry	OSUC Orléans	OZCAR-Tourbières
André	Francois	CNRS/ OMP, Toulouse	OZCAR-Sedoo: M-TROPICS/ BVET, MSEC
Kouadio	Jules	Univ. Gustave Eiffel/ OSUNA Nantes	OZCAR-Observil

Table 4: Theia/ OZCAR observatories data manager(s)- Former collaborators

Name	**Forename**	**Employer/ UMR, town**	**Observatory data base**
Alexandre	Hugues	/ Ecolab, Toulouse	OZCAR-AURADE – SI ECOLEAU
Decoupes	Rémy	Univ La Réunion/ OSU Réunion, La Réunion	OZCAR-ERORUN
Gabarrot	Franck	CNRS/ OSU Réunion, La Réunion	OZCAR-ERORUN
Gautier	Patricia	ENIGMA/ Geosciences, Rennes	OZCAR-H+
Guihéneuf	Nicolas	/ Geosciences, Rennes	OZCAR-H+
Cochonneau	Gerard	IRD/ GET, Toulouse	OZCAR-HYBAM
Silvera	Norbert	IRD/ IEEES, Paris	OZCAR-M-TROPICS/ MSEC
Gogo	Sébastien	CNAP/ ECOBIO Rennes/ ISTO, Orléans	OZCAR-Tourbieres
Paroissien	Jean-Baptiste	CNRS/ ISTO Orléans	OZCAR-Tourbières

- Indicate who is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised.

Sylvie Galle (IGE) is responsible for implementing the DMP, and for ensuring it is reviewed and, if necessary, revised.
Contact: sylvie.galle@ird.fr

- Consider regular updates of the DMP.

The DMP will be reviewed at least every two years and revised as needed.

6b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

Data Terra is a research infrastructure project dedicated to Earth system observation data. Created in 2016, it falls within the Ministry for Higher Education, Research and Innovation (MESRI) 2019-2021 national roadmap. The fact that Data Terra is an IR project on the French national roadmap in principle guarantees that the French state, through supervisory authorities, is committed to the sustainability of human resources, operations, and investment. It is the role/purpose of the Data Terra IR to guarantee the means for the sustainability of an IS such as Theia/OZCAR.

The Gaia Data PIA3 Equipex+ project led by Data Terra and its partners (16.2 M€, 2021-2028) aims to fund a GlobAl IntegrAted DATA and services research infrastructure for the Earth system, biodiversity and environment observation, modelling and understanding. This includes all Theia/OZCAR equipment needs.

- Explain how the necessary resources (for example time) to prepare the data for sharing/preservation (data curation) have been costed in. Carefully consider and justify any resources needed to deliver the data.These may include storage costs, hardware, staff time, costs of preparing data for deposit, and repository charges.
- Indicate whether additional resources will be needed to prepare data for deposit or to meet any charges from data repositories. If yes, explain how much is needed and how such costs will be covered

It is the role of the Theia/OZCAR IT team, with the contribution of the observatory IT teams, to ensure that the data will be FAIR. Thus, Theia/OZCAR will have to provide technical support for the data producers that do not have IT teams. This technical support is not fully funded at the time of the present DMP.

Funding for this task is partly included in the OneWater Data specific program of the PEPR OneWater that should start in 2023 for 8 years.

7. References

Braud, I., V. Chaffard, C. Coussot, S. Galle, P. Juen, and the Theia/OZCAR team (25 persons). « Building the Information System of the French Critical Zone Observatories Network: Theia/OZCAR-IS ». Hydrological Sciences Journal Special Issue: Hydrological data: opportunities and barriers (4 mai 2020): 02626667.2020.1764568. doi: 10.1080/02626667.2020.1764568.

Braud, I. Chaffard, V., Coussot C. and Galle S., 2022. Application of the InteroperAble Descriptions of Observable Property Terminologies (I-ADOPT) Research Data Alliance (RDA) Recommendation by the Theia/OZCAR research community, 2021. https://doi.org/10.15497/RDA00071

Chaffard, V., Coussot, C., Galle, S., Braud, I., and Cailletaud, R.: Mise en oeuvre des principes FAIR pour les donnees de recherche des surfaces continentales: enjeux, freins et bénéfices. Exemple du système d'information Theia/OZCAR., Colloque international "Gestion ouverture des données de la recherche : panorama et perspectives en Afrique," Cotonou (Bénin), 2022.

Coussot, C., Chaffard, V., Braud, I., Galle, S., and Cailletaud, R.: Le thésaurus Theia/OZCAR : un service terminologique pour faciliter la découverte, l'interopérabilité et la réutilisation des données du système d'information sur les surfaces continentales Theia/OZCAR., Colloque international "Gestion ouverture des données de la recherche : panorama et perspectives en Afrique," Cotonou (Bénin), 2022.

Boudevillain, B., G. Delrieu, A. Wijbrans and A. Confoland. 2016. A high-resolution rainfall re-analysis based on radar–raingauge merging in the Cévennes-Vivarais region, France. Flash floods, hydro-geomorphic response and risk management 541: 14-23. doi: 10.1016/j.jhydrol.2016.03.058.

Gaillardet, J., I. Braud, F. Hankard, and the OZCAR team (146 persons). « OZCAR: The French Network of Critical Zone Observatories ».Vadose Zone Journal 17, no 1 (2018): 0. doi: 10.2136/vzj2018.04.0067.

Pellarin, T., T. Tran, J.-M. Cohard, S. Galle, J.-P. Laurent, P. de Rosnay, et T. Vischel. « Soil moisture mapping over West Africa with a 30-min temporal resolution using AMSR-E observations and a satellite-based rainfall product ». Hydrology and Earth System Sciences13, no 10 (15 octobre 2009): 1887-96. doi: 10.5194/hess-13-1887-2009, 2009.

Suominen, O., Ylikotila, H., Pessala, S., Lappalainen, M., Frosterus, M., Tuominen, J., Baker, T., Caracciolo, C., Retterath, A. (2015). Publishing SKOS vocabularies with Skosmos. Manuscript submitted for review, June 2015.

Theia/OZCAR (2018): Theia/OZCAR thesaurus. IRD, INRAE, CNRS-INSU, OSUG. doi: 10.17178/67b5a1d5-8c8c-4a94-a646-1cca1d0adf79

Vischel, T., G. Quantin, T. Lebel, J. Viarre, M. Gosset, F. Cazenave, et G. Panthou. « Generation of High-Resolution Rain Fields in West Africa: Evaluation of Dynamic Interpolation Methods ». Journal of Hydrometeorology 12, no 6 (décembre 2011): 1465-82. doi: 10.1175/JHM-D-10-05015.1.

---

[i] FairTOIS ANR project: https://anr.fr/fr/lanr-et-la-recherche/engagements-et-valeurs/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-fairtois/

[ii] Theia/OZCAR Information System: https://www.theia-land.fr/product/donnees-in-situ/

[iii] OZCAR Research Infrastructure: https://www.ozcar-ri.org/

[iv] FAIR principles : https://www.go-fair.org/fair-principles/

[v] Data Terra Research Infrastructure: https://www.data-terra.org/

[vi] Theia/OZCAR pivot data model: https://github.com/theia-ozcar-is/data-model-documentation

[vii] ISO 19115/INSPIRE standard: https://www.iso.org/standard/53798.html

[viii] DataCite: https://datacite.org/

[ix] OGC catalog service: https://www.ogc.org/standards/cat

[x] GCMD keywords: https://earthdata.nasa.gov/about/gcmd/global-change-master-directory-gcmd-keywords

[xi] SANDRE : Service d'Administration National des Données et Référentiels sur l'Eau, http://id.eaufrance.fr/gpr/41

[xii] ScanR: https://scanr.enseignementsup-recherche.gouv.fr/

[xiii] ISO 19115 topic categories standard: https://inspire.ec.europa.eu/metadata-codelist/TopicCategory

[xiv] INSPIRE Spatial Data Themes: https://www.eionet.europa.eu/gemet/en/inspire-themes/

[xv] Skos: https://www.w3.org/2004/02/skos/

[xvi] Theia/OZCAR thesaurus: https://in-situ.theia-land.fr/skosmos/theia_ozcar_thesaurus/en/

[xvii] Skosmos : http://skosmos.org/

[xviii] Schema standards: https://www.schema.org/

[xix] DCAT standards: https://www.w3.org/TR/vocab-dcat-2/

[xx] Theia/OZCARdata model documentation: https://github.com/theia-ozcar-is/data-model-documentation/tree/master/standard-mapping

[xxi] Theia/OZCAR GitHub repository: https://github.com/theia-ozcar-is/data-model-documentation/tree/master/pivot-data-model

[xxii] Theia/OZCAR web portal: https://in-situ.theia-land.fr/

[xxiii] Data citations principles: https://www.force11.org/datacitationprinciples

[xxiv] see for example the common GDPR guidelines of Paris Lumières, Paris 8 and Paris Nanterre universities (in french): https://www.u-plum.fr/guide-reglement-general-pour-la-protection-des-donnees/

[xxv] Theia land surface data pole: https://www.theia-land.fr/product/donnees-in-situ/

[xxvi] French Water Framework Directive: https://www.ecologie.gouv.fr/gestion-leau-en-france

[xxvii] ISO 19115 responsible party role standard: https://inspire.ec.europa.eu/metadata-codelist/ResponsiblePartyRole

[xxviii] ISO 8601 date time standard: https://www.iso.org/standard/70908.html

[xxix] ISO 3166 Country code standard: https://www.iso.org/iso-3166-country-codes.html

[xxx] INSPIRE D2.9: Guidelines for the use of Observations & Measurements and Sensor Web Enablement-related standards in INSPIRE (2016, 84 p): https://inspire.ec.europa.eu/file/1638/download?token=EtGplOtQ

[xxxi] Complex Property Model (CPM) Ontology: http://adamml.github.io/opm-owl/

APPENDIX 1

*Table 4: List of OZCAR observatories (data producers) and their data centres**

OZCAR Observatory	Data Centre name	Data Centre Type*	Location (city)	Notes

ADES	BRGM	Institutional	Orléans	
H+	OSUR	OSU	Rennes	
ObseRA	IPGP	OSU	Paris	
Tourbières	OSUC	OSU	Orléans	
OHGE	EOST	OSU	Strasbourg	Data base being built using the Open BDOH solution
M-Tropics HYBAM	OMP-sedoo	OSU	Toulouse	
ERORUN	OSU-Réunion	OSU	La Réunion	
Karst	OREME	OSU	Montpellier	
Auradé	Ecolab	UMR	Toulouse	Data will be part of OMP OSU SEDOO Information System
OSR	CESBIO	UMR	Toulouse	Manages all CESBIO data, in link with the OMP OSU
AMMA-CATCH OHM-CV Cryobs-Clim	IGE	UMR	Grenoble	Joint management with the OSUG OSU
OMERE	LISAH	UMR	Montpellier	
Agrhys	SAS	UMR	Rennes	developed in relation with OSUR OSU
Draix-Bléone GIS Oracle OTHU/Yzeron Real Collobrier OHM-CV	RiverLy	UMR	Lyon	« BDOH » Database developed at UR RiverLy and used by over laboratories
Observil	OSUNA	OSU	Nantes	A data base is being built for this SNO

*: There are three data centre types

- Institutional: data centre operated by a research institution.
- OSU: “Observatoire des Sciences de l’Univers” are local federative structures in charge of observation and data management.
- UMR: “Unité Mixte de Recherche”. This means that the data are managed either by the UMR data centre or by the researchers themselves.