Implementing FAIR principles for dissemination of data from the French OZCAR Critical Zone Observatory network:
the Theia/OZCAR Information System

Isabelle Braud\textsuperscript{(1)}, Véronique Chaffard\textsuperscript{(2)}, Charly Coussot\textsuperscript{(2,3)}, Sylvie Galle\textsuperscript{(2)}, Rémi Cailletaud\textsuperscript{(3)} and the Theia/OZCAR IT Team

\textsuperscript{(1)} INRAE, RiverLy, Lyon, France
\textsuperscript{(2)} IGE, CNRS-IDR-UGA, Grenoble, France
\textsuperscript{(3)} OSUG, CNRS-IDR-UGA-INRAE, Grenoble, France
Outline of the presentation

1. The OZCAR Critical Zone Network and objectives of the Theia/OZCAR IS
2. Collecting information on data management and future users needs
3. Design of the Theia/OZCAR Information System
4. Implementation of the data discovery portal
5. Conclusions and perspectives

On each slide, click on the arrow to come back to the presentation outline
A network of 21 observatories

- Documents ~ 60 sites
- In France and in Southern countries

A long history of observation
- Observatories developed independently
- Heterogeneity in data management

Diversity of the objects of interest and observations
- Watersheds, rivers
- Aquifers
- Glaciers
- Permafrost
More than 300 measured variables

- **Point** time series: meteorology, hydrology, hydrogeology, glaciology, surface energy balance, sediment fluxes, geochemical elements and contaminant concentrations
- **Soil cores**
- **2D** geophysical profiles
- **Maps** (raster or vector) characterizing the sites: land use, DTM, soil physical properties
- **Surveys**: crop rotations

=> A large diversity in collected variables and names
**Objectives of Theia/OZCAR IS**

- **A unique data portal** to access **transparently** in situ data documenting continental surfaces and the critical zone, that are presently scattered in various information systems.

- Respect the accessibility and interoperability principles in relation with the European INSPIRE directive and **FAIR principles**

- Foster **DOI declaration** on data sets

- Offer **services and interoperability** with other portals, in particular the Theia remote sensing portal, Data Terra Research Infrastructure and European Research Infrastructures (e.g. European Long Term Ecological Research – eLTER- RI)

- Design the IS using data from OZCAR-RI that is representative of the diversity of in situ data describing continental surfaces, and then extend the IS to other RIs, data from projects or sites for the calibration/validation of satellite products.
Methodology

• A “Tour de France” of the observatories
  ➢ to understand how data management is organized
  ➢ identity human resources and potential contacts,
  ➢ Collect expectations and fears regarding the project

• Organization of working groups at the OZCAR 2018 annual meeting to collect users expectations with regards to the web interface
  ➢ Criteria for data search for the scientist user
  ➢ Criteria in relation with data provision and statistics about their use for data producers

• Participation in the InterPole working groups to share ideas and practices with the other data poles
Which data management?

- Observatories generally related to regional data centers (Science of the Universe Observatories) or to institutional data management
- Each observatory has its own data management system and dissemination
- A large heterogeneity in data discovery and access
Different existing data portals or file repositories
Different existing data portals or file repositories

• A large diversity in data management with file repositories, metadata portals, data portals fully interoperable or not, or still under construction

• A substantial effort to organize, sometimes harmonize data in the various observatories

⇒ Given the human power invested, necessity to make the best use of existing organizations
Outcome of the future users and data providers consultation

Search criteria

- Variable names (normalized)
- Feature of interest (catchment, river, etc.)
- Type of climate, geology
- Institutions (observatories, funders, projects, etc...)

Which expectations?

- Metadata AND data in the same output formats
- Data quality and documentation: ensured by data producers
- Must be first useful for data producers themselves (statistics of downloading, end up to be the best place to download their own data)

Human resources

- IT skills available but scattered in different locations
- Wish to keep the data close to producers to ensure data quality
- IR skills not available everywhere to build similar data management systems

=> Necessity to make the best use of existing systems and organize information fluxes between local systems and the central system
Principles for building the Theia/OZCAR IS

User-oriented approach (data producers or not)

- Intuitive data search (dynamic map)
- Standard vocabulary (variables names and categories)
- Export: standard formats

Update in real time

- data flow: observatories local IS => Theia / OZCAR central IS
- Common metadata model: pivot format (Braud et al., HSJ, in press)

Association of researchers / IT team; round trip between data producers and project team; use of the Agile approach
What has been done so far for data discovery?

1. Building of a **controlled vocabulary** (variables names and categories)

2. Definition of the required **metadata**: analysis of standards (ISO19115, INSPIRE, DataCite, O&M, etc..) to define the information flux to be organized between observatories and the central Theia/OZCAR IS

3. Definition of a **pivot data model** for exchanging the information
   
   https://github.com/theia-ozcar-is/data-model-documentation

4. Definition of the architecture of the Information System and building of a prototype web portal
Hierarchized controlled vocabulary

- Publication of the thesaurus (Linked Open data)
- Semantic links with international thesauri

Building of data fluxes:

1) Different data producers, different formats

2) The pivot data model allows (i) information collection; (ii) to update them in real time

3) The Theia/OZCAR IS is able to answer requests from humans and machines
ISO 19115 / Inspire

https://github.com/theia-ozcar-is/data-model-documentation
See also slides 26 and 27
**Web portal for data discovery**

**January 2020: beta version is online**
- Search by facets (variables, geography, observatories, funders, etc.)
- Metadata only
- Observations from 7/21 observatories are visible

**Future steps**
- Validate the ergonomics of the portal and its functionalities
- Interfacing ALL OZCAR observatories (setting up data flows)
- Enable data search on features of interest
- Set up user authentication
- Allow data downloading in a common format (.csv or NetCDF)
- Allow interoperability (harvesting by machines)
A network of 21 observatories

- Documents ~ 60 sites
- In France and in Southern countries

In 2020:

a third of the observatories have set up a data flux with Theia/OZCAR IS

A long history of observation
- Observatories developed independently
- Heterogeneity in data management

Diversity of the objects of Interest and observations
- Watersheds, rivers
- Aquifers
- Glaciers
- Permafrost
Data portal

Search by facet
- by geographic area
- by time period
- by variable
- by observatory
- By guardianship

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

Quick overview:
station name, producer, measured variables
Portail de données

Detailed information:
- list of variables, station names, producer, other variables in the dataset
Conclusions and perspectives

January 2020: beta version is online

• Search by facets (variables, geography, observatories, funders, etc.)
• Metadata only
• Observations from 7/21 observatories are visible

Future steps

• Validate the ergonomics of the portal and its functionalities
• Interfacing ALL OZCAR observatories (setting up data flows)
• Enable data search on features of interest
• Set up user authentication
• Allow data downloading in a common format (.csv or NetCDF)
• Allow interoperability (harvesting by machines)

Genericity of the approach

• A method that can be used for other distributed existing data infrastructure
• Open solutions that can be reused by others
To learn more about the project:

To access the portal, the thesaurus and the project Github
https://in-situ.theia-land.fr/
https://github.com/theia-ozcar-is

Contacts:
Isabelle.braud@inrae.fr
Sylvie.galle@ird.fr
Veronique.chaffard@ird.fr
Acknowledgements:
The authors thank all the participants of the “Tour de France” meetings, and those that contributed to the working groups during the OZCAR Fréjus meeting in April 2018.
The authors are grateful to the Theia technical team (N. Baghdadi, A. Sellé, S. Debard) and more generally the interpole technical group led by F. Genova, including Odatis, Form@terre, Aeris, Theia and PNDB.
The persons involved in the Trajectories project from Grenoble Alpes University, and in the Zone Atelier Network are also thanked for fruitful exchanges.
The work is conducted within the OZCAR-RI, supported by the French Ministry of Research, French research institutions, universities and the French Agence Nationale de la Recherche (ANR, project FariTOIS ANR-19-DATA_0003). The work is also part of the Theia land data pole of the Data Terra Research Infrastructure.
CNRS/INSU and IRD are also thanked for funding the work of the third author.
A common information system: Theia/OZCAR

Fundings: INSU, IRD, IR OZCAR, ANR FairTOIS

- Open data
- Thematic data pole (continental surface)
- 4 Thematic data pole (5?)

- Global change / Critical zone
- Network: 21 long term in situ observatories

EOSC Pillar: WP on « uses cases »
PHIDIAS: on demand service

« European Long Term Ecosystem Research » RI on the European road map since 2018. (Portal to build)
Diagramme de classe

Pivot data model (1/2)