Building the Information System of the OZCAR French Critical Zone Observatories network: principles and first prototype
Isabelle Braud, C. Coussot, V. Chaffard, S. Galle, P. Juen

To cite this version:

HAL Id: hal-02609282
https://hal.inrae.fr/hal-02609282
Submitted on 16 May 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
1. Context and objectives

Context: the OZCAR Research Infrastructure

- OZCAR: a network of 22 observatories monitoring various compartments of the critical zone all over the world (about 50 sites, Fig. 1) (Gaillardet et al., 2018).
- Data presently scattered in various portals or data repositories.
- A large diversity of observed variables: time series, gridded and vector 2D data.
- Water, energy, sediment and matter transport, geochemistry in catchments.
- Transport of water, solutes and reactive elements in groundwater, fractured or karstic aquifers.
- Glaciers, snow and permafrost processes.
- Soil profiles/carrots and sampling specimens.
- Geophysical data (2D).
- Vector GIS data.
- Surveys (agricultural practices).
- Raster data and remote sensing products: Land use, land cover, high resolution DTM.

Objectives of the Theia-OZCAR Information System (IS)

- Theia: one of the French data pools dedicated to continental surfaces, exposing presently mainly remote sensing data.
- A need to extend the Information System (IS) to in situ data (Galle et al., 2018).

Objectives of the project

- Propose a unique data portal to access in situ data documenting continental surfaces and the critical zone, that are presently scattered in various information systems.
- Keep the data close to data producers and make the best use of existing data management systems.
- Define information fluxes between observatories and the Theia/OZCAR IS (distributed architecture).
- Be useful to observatories managers (identification of data users, publications using data, putting DOI on data sets, etc...).
- Offer services and interoperability with other portals.
- Foster data discovery and exploration, their sharing and reusability, their citation (Fig. 2).

2. Methodology

- System uses cases:
  - User-oriented approach: dialog between scientists & IT teams of the source observatories (working groups, meetings).
  - Story mapping (AGILE method).

Building strategy:

- Definition of common controlled vocabulary based on domain-specific thesaurus.
- Open source softwares.
- Definition of common metadata based on metadata standards.
- Put in practice FAIR principles.
- Organization of data fluxes between data producer IS and Theia/OZCAR-IS using a pivot data model (Fig. 3).

3. Implementation

Choice of a common Theia-OZCAR hierarchized vocabulary (Fig. 4)

- Controlled vocabulary based on hierarchized concepts of GCMD Earth Science keywords thesaurus and Sandre thesaurus for water chemistry.
- Harmonized variables names and categories.
- From about 3000 data producers names to about 250 controlled names.
- Publication in Skos using SKOSMOS.

Definition of a pivot model to build information flux (Fig. 5)

- Concepts and information based on metadata/data exchange standards (ISO19115/Inspire, DataCite, O&M) and some system functional requirements.

4. System architecture and prototype of the web interface

- Building the Theia/OZCAR Information System prototype: the web interface.

5. Next steps

- Continue the implementation of the information flux between the observatories and the Theia/OZCAR IS (5 observatories included up to now).
- Continue the implementation of search functionalities on the web interface: adding search based on variable name thesaurus.
- Test if the web interface is user-friendly with future users.
- Put the system in production for metadata (end of 2019) and extend it to data (2020).