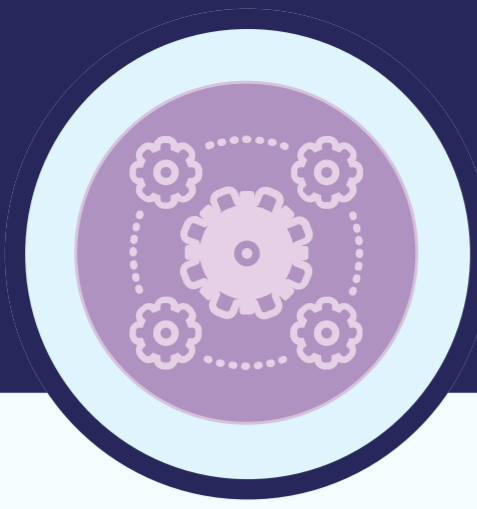




Interoperability



Metadata & Ontologies



Life Science



Data Terra

Enhance the semantic functionality on the national Earth & Environmental Data Repository by integrating it with the EarthPortal

Overview

Easy Data is the French repository for long-tail data relating to the Earth and the Environment. EaSy Data uses vocabularies initiated by scientists to fill in specific metadata elements such as topics or keywords. These vocabularies are incomplete and do not reflect the complex diversity of data deposited in EaSyData. Other vocabularies exist and could be used to complete this vocabulary.

The aim of this use case is to improve the semantic functionality of EaSyData by connecting it to the EarthPortal. By combining the vocabularies displayed in the EarthPortal and EaSy Data vocabulary with associated services, we can offer researchers a more comprehensive set of tools to specify metadata and thus improving the semantic capabilities of EaSyData.

Context and objectives

EaSy Data is a French national data repository dedicated to long-tail environmental and terrestrial data, using the ISO 19115 standard to describe them. EaSy Data uses Geonetwork to store metadata and a custom application layer has been developed to support the data repository and related functions (repository, search).

Community-controlled vocabularies are used to fill metadata elements related to keywords or topics. These specific vocabularies have been defined by scientists and are maintained in an experimental registry based on UKGovLD that has limited facilities to administer and suggest new keywords.

Furthermore, these vocabularies are incomplete and do not reflect the complex diversity of data deposited in EaSyData. Other vocabularies exist and could be used to complete this vocabulary. For example, some of the thesauri produced by French data clusters, such as those supplied by Theia/Ozcar, or by European infrastructures, such as those defined by EPOS or ACTRIS, could be used.

Contributors



Christelle Pierkot
(Data Terra)



Guillaume Alviset
(Data Terra)



Hélène Bressan
(BRGM)



Interoperability



Metadata & Ontologies



Life Science



Data Terra

Challenges and solutions

In EaSy Data, we need to use community vocabularies at various levels:

- **administration**, to manage users by topic, e.g. moderators linked to specific topics
- **depositors**, to better describe datasets and avoid spelling errors, etc.
- **searchers**, to improve search guidance.

The use of a catalog to reference and update existing vocabularies, and potentially to host specific vocabularies for EaSyData, adds significant value.

To achieve this goal, we need to change the vocabulary tool used in EaSyData to better manage vocabularies and benefit from enhanced services.

The EarthPortal is a thematic semantic artefact catalog and repository for the Earth sciences using the OntoPortal technology. EarthPortal provides tools useful to improve the EaSy Data semantic functionality, such as:

- **the Annotator** (makes term suggestions based on text input)
- **the Recommender** (suggests relevant SA based on text input)
- **Mappings** (generates, stores and displays mappings between SA)

Data Terra's thesauri, as well as vocabularies produced by European infrastructures and other more generic but commonly used SA are already available in EarthPortal. EaSy Data will harvest these vocabularies directly from the Earth Portal's existing REST API, to offer users more terms than those initially defined and to use the annotation service offered by EarthPortal. This will improve the user experience in two ways:

- **To populate keywords and topics metadata**, the EarthPortal will allow the user to check additional vocabularies to those currently used by EaSy Data.
- **Annotator service will be used to suggest new terms** from the user-written abstract to enrich the metadata with new terms.

Expected/Measured Impacts

From a user perspective:

- **A better access to the vocabularies** and the possibility to contribute to them;
- **An extended semantic description of the datasets**, allowing better discovery of related resources;
- **Semantically enriched metadata of the datasets** by looking at the related SA to suggest related vocabulary concepts.

From an administrator's perspective:

- **a better management of the EaSy Data vocabularies** and a smarter use of other vocabularies.

Contributors



Christelle Pierkot
(Data Terra)



Guillaume Alviset
(Data Terra)



Hélène Bressan
(BRGM)