



**HAL**  
open science

# Enhancing FAIRness of Spatiotemporal Data through Semantics: A Case Study with EarthPortal and EaSy Data

Christelle Pierkot, H el ene Bressan, Guillaume Alviset

► **To cite this version:**

Christelle Pierkot, H el ene Bressan, Guillaume Alviset. Enhancing FAIRness of Spatiotemporal Data through Semantics: A Case Study with EarthPortal and EaSy Data. SIMbig 2024 - 11th International Conference on Information Management and Big Data, Nov 2024, Ilo, Peru. hal-04908547

**HAL Id: hal-04908547**

**<https://hal.science/hal-04908547v1>**

Submitted on 23 Jan 2025

**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 destin ee au d ep ot et  a la diffusion de documents scientifiques de niveau recherche, publi es ou non,  emanant des  tablissements d'enseignement et de recherche fran ais ou  trangers, des laboratoires publics ou priv es.



Distributed under a Creative Commons Attribution 4.0 International License

# Enhancing FAIRness of Spatiotemporal Data through Semantics: A Case Study with EarthPortal and EaSy Data

Christelle Pierkot<sup>1</sup>[0000-0002-2591-3311], Helene Bressan<sup>2</sup>[0000-0001-5673-4474], and Guillaume Alviset<sup>1</sup>[0009-0004-4295-6593]

CNRS, UAR Data Terra, MTD, 500 rue JF Breton, 34090 Montpellier, France  
{christelle.pierkot,guillaume.alviset}@cnrs.fr  
BRGM  
{h.bressan@brgm.fr}

**Abstract.** The study of the Earth system involves many disciplines working together to address complex issues such as climate change. In France, Data Terra, the national research infrastructure related to Earth System, provides data discovery, access and dissemination services to support researchers in their work. Within this framework, the EaSy Data repository facilitates the sharing and enhancement of research data, while EarthPortal, a catalogue of FAIR semantic artefacts, promotes the use and dissemination of controlled vocabularies and ontologies. The adoption of the FAIR principles has become widespread among Earth system data managers, with several initiatives focused on advancing these principles. Interoperability, in particular, remains a critical yet challenging aspect, essential for ensuring that data can be effectively combined, interpreted, and used by both humans and machines. To fulfill the FAIR criteria for interoperability, it is imperative that data and metadata employ controlled vocabularies and structured representation models that are consistent with these principles. This article demonstrates how the integration of EarthPortal within the EaSy Data repository, significantly improves data interoperability and optimizes the user experience.

**Keywords:** Semantic artefacts · Spatiotemporal data · FAIR principles · Interoperability · Data Repository · EarthPortal · EaSy Data

## 1 Introduction

The study of the Earth system encompasses numerous disciplines that increasingly collaborate to address complex issues such as climate change. In France, the national research infrastructure, Data Terra [1], is responsible for managing data related to five key components of the Earth system: the Ocean, the Atmosphere, the Solid Earth, the Continental Surface, and Biodiversity. In addition to distributed infrastructure services, Data Terra offers data discovery, access, and dissemination services, thereby facilitating researchers' work. Among these services, the national data repository, EaSy Data [2], allows researchers to share their data for dissemination and use by other potential users.

The FAIR principles, established by [3] to optimize data management in repositories, are based on four key concepts: Findability, Accessibility, Interoperability, and Reusability. These principles are widely adopted by Earth system data managers, and various initiatives and projects have been launched to enhance them. In particular, adherence to the principle of Interoperability in data repositories ensures that the data is downloadable, usable, and understandable with other data,

both by humans and machines. However, experience shows that Interoperability is often the most challenging principle to achieve when it comes to sharing and reusing data.

The use of semantics in data repositories can significantly improve Interoperability outcomes. The FAIR principles' I1 and I2 criteria specify that (i) data descriptions must follow a controlled vocabulary and a representation model to structure this vocabulary, and (ii) data and metadata must use vocabularies that comply with the FAIR principles. These vocabularies, documented and accessible via a persistent identifier (PID), must be machine-readable. Therefore, a thoughtful application of semantics in data repositories is essential to strengthen their Interoperability. In this article, we illustrate, through an application example in the EaSy Data repository, how the use of EarthPortal [4], a catalog/repository of semantic artefacts, can not only improve the Interoperability of data repositories but also simplify the deposit experience for users.

The paper is structured as follows: First, based on the recommendations of the FAIR principles, we analyse how semantics could be used to improve the interoperability of data repositories. Next, we present EaSy Data, a thematic data repository, and the EarthPortal, a catalogue of semantic artefacts, and outline the various possible uses of combining the two. We then present the example of the connection between EarthPortal and EaSy Data and the added value this brings. Finally, we propose some future perspectives for this work.

## 2 Unlocking Interoperability in Data Repositories: The Role of FAIR Principles and Semantic Tools

The FAIR principles have played a crucial role in establishing best practices for data management within scientific communities, gaining significant support from research funding bodies to encourage their adoption. In France, the ANR Science Ouverte research program has supported various projects centered around these principles, such as FairTOIS [8], Semantics4Fair [9], BRIDGE [10], and Copilote [7]. The European Open Science Cloud [11] program is also pivotal in implementing these principles, supporting major projects like FAIRsFAIR [12], FAIR-IMPACT [13], ENVRI-FAIR [14], and FAIR-EASE [5], and forming working groups on FAIR data, such as the FAIR Metrics and Data Quality Task Force [22] (2021-2023) and the FAIR Metrics and Digital Objects Task Force (2024-). Internationally, the Research Data Alliance [15] is distinguished by its ongoing commitment to initiatives related to the FAIR principles. Several working groups have emerged in recent years (such as FAIR4RS, FAIR4ML, FAIR4VRE, and FAIR mappings), demonstrating the growing interest of the scientific community, including Earth sciences, in integrating these principles into their respective fields.

Compliance with the FAIR principles in data repositories ensures that data is Findable, Accessible, Interoperable, Reusable, and can be integrated with other datasets. The interoperability criteria of the FAIR principles, which stipulate that (I1) data descriptions must follow a controlled vocabulary and a representation model to structure this vocabulary, and (I2) data and metadata must use vocabularies compliant with the FAIR principles, are essential for the implementation of data repositories, ensuring that data is comprehensible and usable by both humans and machines. However, as highlighted by [16], the sub-principles under Interoperability are often considered the most challenging to achieve and the most critical for meeting FAIR requirements. To meet criterion I1, metadata that explicitly describes data in repositories must follow a standardized, common, shared, and semantically rich schema. In data repositories, semantic artefacts (a generic term proposed by the FAIRsFAIR project to denote semantic objects such as vocabularies, ontologies, etc. [17]) have become essential to fulfilling the interoperability criteria of the FAIR principles.

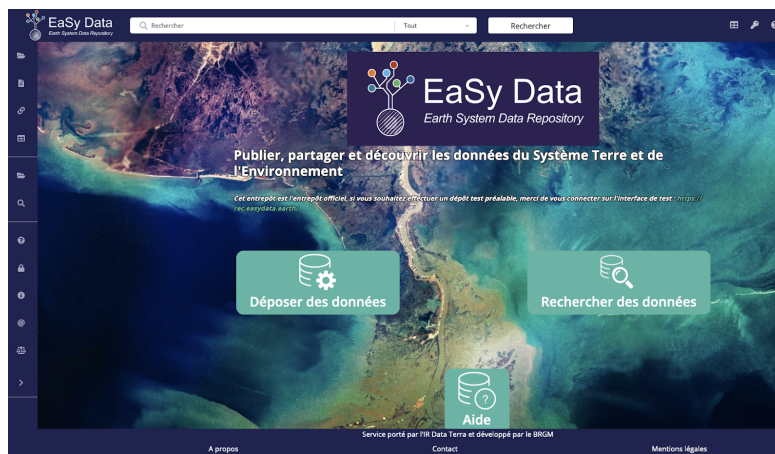
In the field of Earth sciences, various semantic artefacts (SA), such as vocabularies and ontologies, are employed to represent and annotate data in a standardized manner. For instance, the SWEET ontology (Semantic Web for Earth and Environment Technology Ontology) and the GCMD (Global Change Master Directory) thesauri are among the major references. However, these high-level semantic artefacts do not exhaustively cover all the characteristics of the Earth system. Consequently, initiatives have been established to develop more specifically tailored semantic artefacts. In France, the Data Terra research infrastructure, which covers five components of the Earth System, has specified vocabularies for some of these components for use in their data catalogs. The EaSy Data Thesaurus, a unified vocabulary, was created for use in the national thematic data repository, EaSy Data, which hosts research data from Earth system components. At the European level, the Actris infrastructure, dedicated to climate and air quality research, has developed a controlled vocabulary that it uses in its catalog. Similarly, EPOS, the infrastructure for the Solid Earth component, uses a vocabulary in the EPOS-MSL (EPOS Thematic Core Services Multi-Scale Laboratories) services.

To address the multiplication of semantic artefacts and ensure criterion I2, [18], [19] propose ten rules to make the semantic artefacts themselves FAIR (e.g. assigning a unique and persistent identifier, adding metadata, registering the semantic artefact in a semantic repository, etc.). [20] recommend the use of catalogues that allow the sharing, dissemination and reuse of semantic artefacts.

Based on these recommendations, we have handled a concrete example to show how a semantic artefact catalogue can improve the interoperability of spatial data repositories.

### 3 Managing interoperability in the EaSy Data repository

EaSy Data is the French national trusted repository dedicated to research data related to Earth and environmental sciences.



**Fig. 1.** EaSy Data main interface.

With the help of EaSy Data, a researcher can deposit data from all areas of the earth science discipline and request a DOI. The approval of the submission is subject to a moderation process carried out by experts from different disciplines. A search interface is also available to enable the data to be easily found by other users.

In accordance with international standards, EaSy Data is based on a standardized metadata schema, utilizing the ISO 19115 standard for data description. Metadata management is carried out using the Geonetwork tool. Additionally, a web application layer has been developed to facilitate data submission and related functionality (such as search and rights management).

To comply with the FAIR principles, particularly those related to interoperability, several metadata elements within EaSy Data incorporate semantic artefacts. These artefacts are exposed in various semantic web formats and are accessible via APIs. While some semantic artefacts are sourced from publicly available online registers, others have been specifically created for the EaSy Data repository. The table below presents the metadata elements that, from the first version of EaSy Data, were initially associated with one or more semantic artefacts.

**Table 1.** metadata elements linked to a semantic artefact in EaSy Data

Metadata element	Metadata element description	Example	SA used
Topics	General topics or fields in which the dataset fits	seismology, atmosphere	EaSyData_Thematique
Keywords	Keywords describing the specific characteristics of the dataset.	Surface water conductivity , seismic signal	EaSyData_Keywords
Organisation	research organizations	CNRS, IRD	ROR
License	data distribution licence	CC-BY-4.0, etalab-2.0	SPDX Licenses
Geographic Description	geographical description of the dataset's spatial extent	France, California	Geonames
Codes	Codelists for description of metadata datasets compliant with ISO 19115	Point of Contact, Custodian (Role Code) - vector, grid (SpatialRepresentationTypeCode)	ISO 19115 Codelists

The vocabularies used for the metadata elements "Topics" and "Keywords" were defined by scientists. However, these vocabularies remain incomplete and do not fully capture the complex diversity of the data deposited in EaSy Data. Other vocabularies exist and could be utilized to supplement this register, such as certain thesauri produced by the national infrastructure Data Terra or by European infrastructures like those defined by EPOS [23] or ACTRIS [24].

## 4 Improving Interoperability by using the EarthPortal

EarthPortal is a catalog of thematic semantic artefacts and a semantic repository intended for Earth sciences, based on OntoPortal technology.

Developed within the EOSC FAIR-IMPACT project, its implementation adheres to national and European recommendations, aiming to enhance the dissemination and reuse of semantic artefacts in an interdisciplinary research context. EarthPortal also ensures the traceability and recognition of

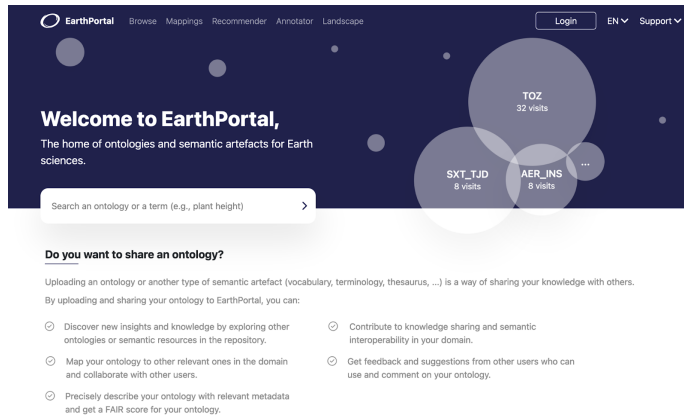


Fig. 2. EarthPortal main interface.

the authorship of vocabularies, thus offering better recognition for their creators. At the European level, the recommendations of the "Semantic Interoperability" Task Force emphasize the importance of recognizing catalogs of semantic artefacts (as described in the interoperability framework of the European Open Science Cloud - EOSC) as a crucial element for the long-term sustainability of any research data infrastructure [21].

EarthPortal hosts semantic artefacts related to Earth and environmental sciences and can be used by external applications via its REST API. These artefacts are both controlled vocabularies in the SKOS format and ontologies specified in OWL. They are organised into categories and groups in the catalogue. Filters can be applied to these elements according to the needs. Thesauri from Data Terra, as well as the vocabularies produced by European infrastructures and other more generic but widely used semantic artefacts (such as SOSA, SWEET, etc.), are actually available in EarthPortal.

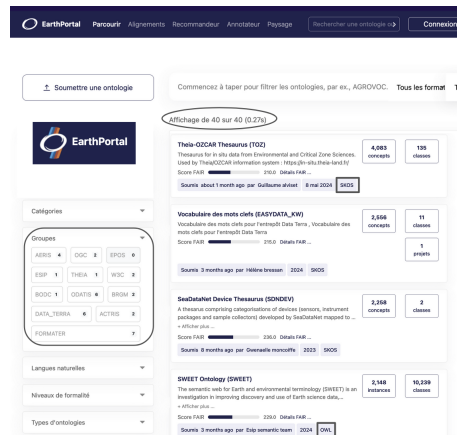


Fig. 3. Semantic Artefacts in EarthPortal

In addition to enabling the dissemination and reuse of Earth system semantic artefacts, EarthPortal offers tools that can be leveraged by third-party applications: the Annotator, which suggests terms based on text or keyword input; the Recommender, which proposes relevant semantic artefacts based on the input text or keywords; and the Mappings tool, which generates, stores, and displays mappings between semantic artefacts.

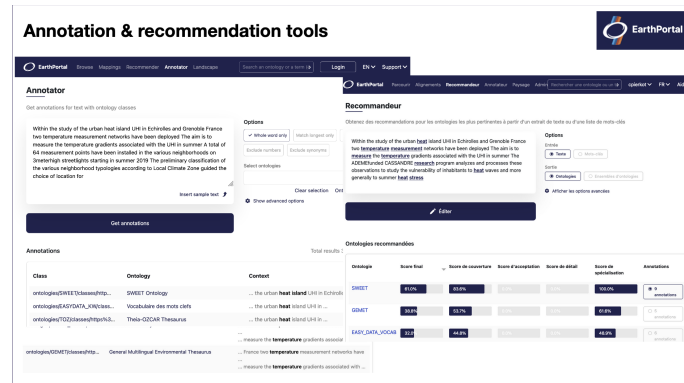


Fig. 4. Annotation and Recommendation services proposed by the EarthPortal

The Annotator is of particular interest as it allows the identification of concepts within the semantic artefacts hosted on EarthPortal, based on keywords or text.

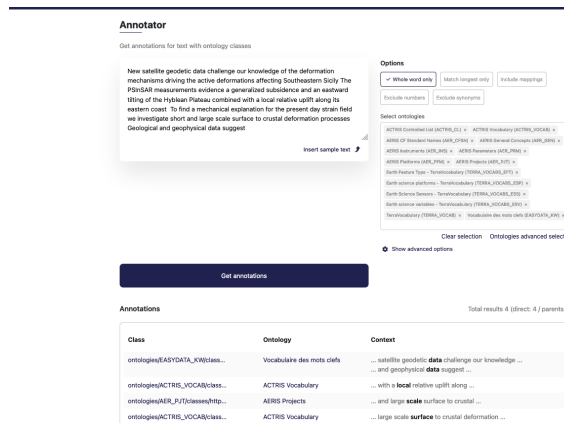
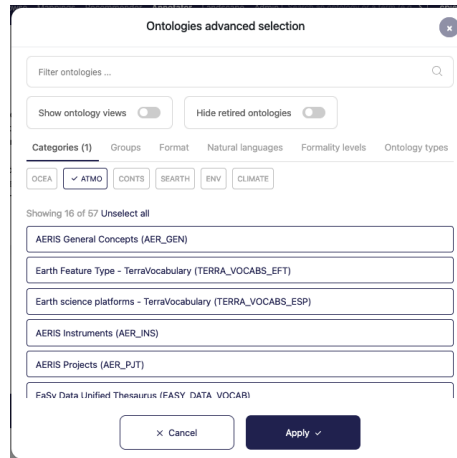


Fig. 5. Annotator tool

Filters can be applied to limit selection to relevant artefacts. In particular, it is possible to filter by category to obtain results specific to a given topic. In the following example, we choose to limit the selection to SA related to the topic Atmosphere



**Fig. 6.** Filtering relevant SA by categories

Thanks to its API, third-party applications such as EaSy Data can connect to EarthPortal to use the semantic artefacts available and exploit the available tools.

## 5 Using EarthPortal within EaSy Data : a concrete application

We have identified three potential use cases for the integration of EarthPortal within EaSy Data:

### – *Use case Data Deposit*

We propose enhancing the process of metadata entry, specifically regarding keywords, through the application of semantic artefacts available in EarthPortal. Currently, within EaSy Data, keywords are defined by thematic experts without a structured hierarchy or precise domain classification. Consequently, these keywords are often incomplete or inappropriate for certain datasets. For instance, keywords related to atmospheric studies may not be suitable for datasets concerning landslides. To address this limitation, we propose connecting EaSy Data to EarthPortal to offer researchers relevant and context-appropriate keywords when depositing datasets.

### – *Use case Search*

The search filters within EaSy Data are currently limited by their reliance on vocabularies specific to the data repository, thereby restricting the breadth of search results. The semantic web, through concept mapping, enables the establishment of relationships between entities across different semantic artefacts. EarthPortal's "mapping" functionality automatically identifies correspondences between vocabularies within its catalog. We intend to leverage this capability to enhance the faceted search in EaSy Data, thereby expanding the scope and relevance of search results.

### – *Use Case Data Harvesting*

- EaSy Data is tasked with harvesting data from other repositories, such as Seanoë, which is exclusively focused on marine data, to make this data accessible on a multidisciplinary scale. However, these repositories often do not employ controlled vocabularies, relying instead on free text to define certain metadata elements, such as keywords. This practice is inconsistent



with EaSy Data’s specifications. To address this issue, we propose utilizing EarthPortal to identify semantic correspondences for these harvested metadata elements, thereby enriching their semantic quality within EaSy Data.

- Furthermore, EaSy Data itself will be harvested by other catalogs, such as the European infrastructure EPOS, which aims to incorporate EaSy Data’s datasets into a broader context. However, these catalogs do not share the same vocabulary as EaSy Data, which complicates data integration. To facilitate the effective utilization of EaSy Data’s metadata within these external catalogs, it is essential to establish correspondences between the respective vocabularies. By using EarthPortal to store the vocabularies of different repositories and leveraging its mapping and annotation features, we can significantly mitigate the risk of information loss.

In the following section, we will focus exclusively on the outcomes of the first use case: data deposit.

## 5.1 Enhancing the Data Deposit experience

When user intends to submit a dataset in EaSy Data, he/she must first provide a title and abstract, and then select relevant topics and keywords from the EaSy Data thesaurus. If the available keywords are not relevant enough, the user can choose to enrich them by using a designated button in the interface.

The screenshot shows the 'Déposer un nouveau jeu de données' (Deposit a new dataset) form in EaSy Data. The form is in English. The title is 'GNSS-TEC perturbation from the 2022 Hunga Tonga volcano eruption over France'. The abstract describes the 2022 Hunga Tonga volcano eruption and the dataset's focus on the first passage of the Lamb wave over France. The repository is 'Ionosphere disturbances from telluric hazards'. The thesaurus includes 'Solid earth' and 'Atmosphere'. The keywords are 'GNSS sensor' and 'Volcano'. A legend indicates that green boxes represent keywords imported from the repository 'Ionosphere disturbances from telluric hazards'. A button labeled 'Suggestion of additional keywords EarthPortal' is highlighted in the 'Mots-clés / Keywords' section.

Fig. 7. Step 1 : Click on the keyword suggestion button

This action triggers some request to the Earth Portal API. The target semantic artefacts are filtered according to the topics selected by the user in EaSy Data, thereby minimizing the risk of misinterpretation between terms outside the desired discipline:

- Call `https://data.earthportal.eu/categories/{Topic}?display=ontologies` where :
  - Topic is the topic given by the user in the EaSy Data form

Earth Portal then processes this abstract through its annotation service to propose semantic concepts.

- Call `https://data.earthportal.eu/annotator?text={Abstract}&ontologies={List of SA}` where :
  - Abstract is the abstract given by the user in the EaSy Data form
  - List of SA is the result of the previous request

The output of the request is a JSON file containing all the correspondences between the terms given in the EaSy Data abstract and the concepts found in the semantic artefacts hosted in the EarthPortal, filtered by category, as shown in the example below :

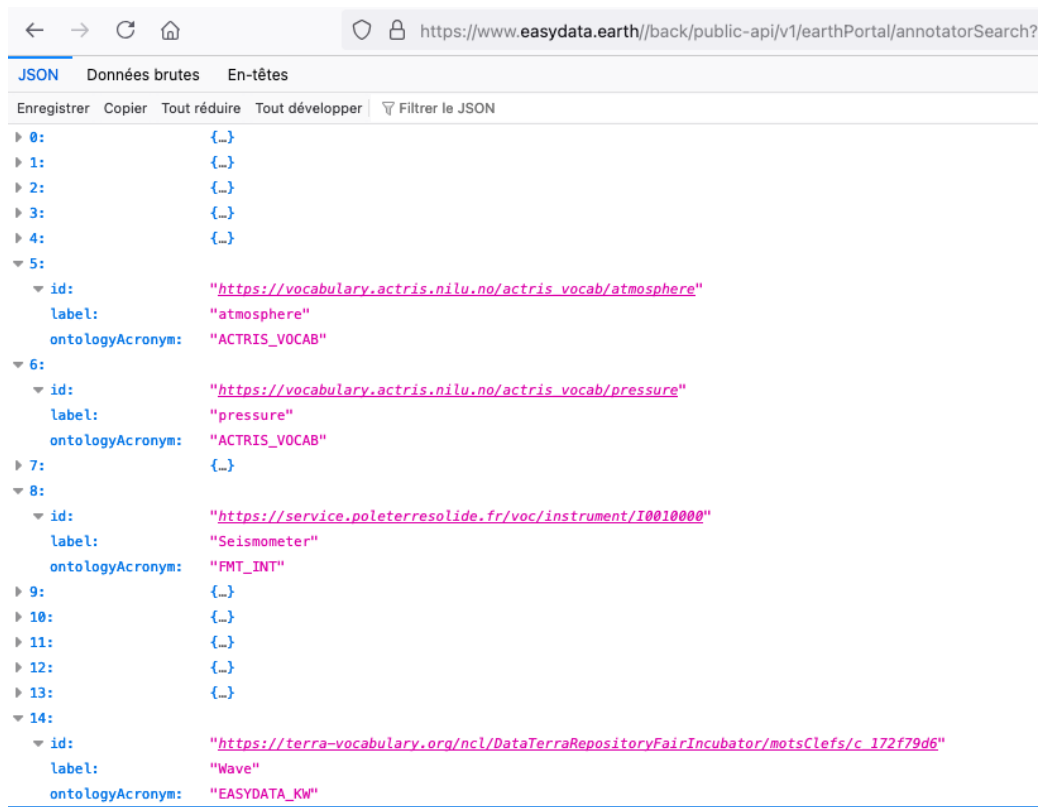


Fig. 8. Part of the JSON file resulting from annotating the above example

EaSy Data interprets this JSON file and suggests the additional keywords in a new window, allowing the user to select those of interest.

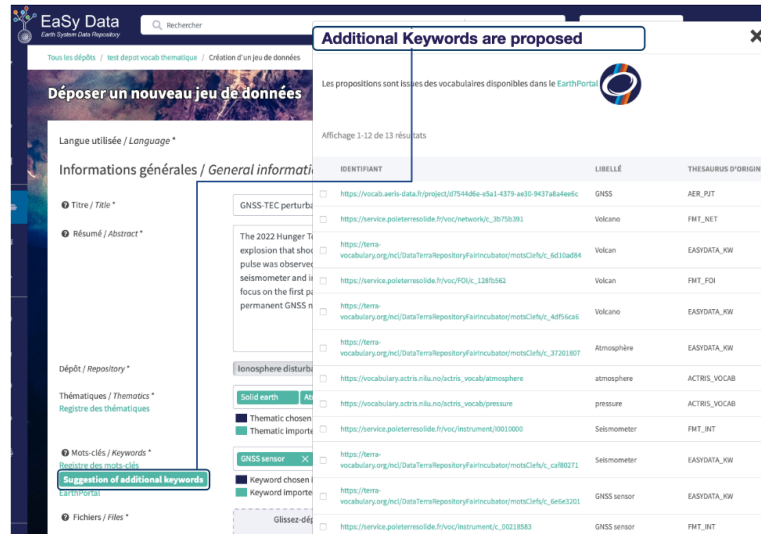


Fig. 9. Step 2 : Suggestion of additional keywords based on the EarthPortal Annotator result

The user selects one or more additional Keyword(s) and send the request to EaSy Data.

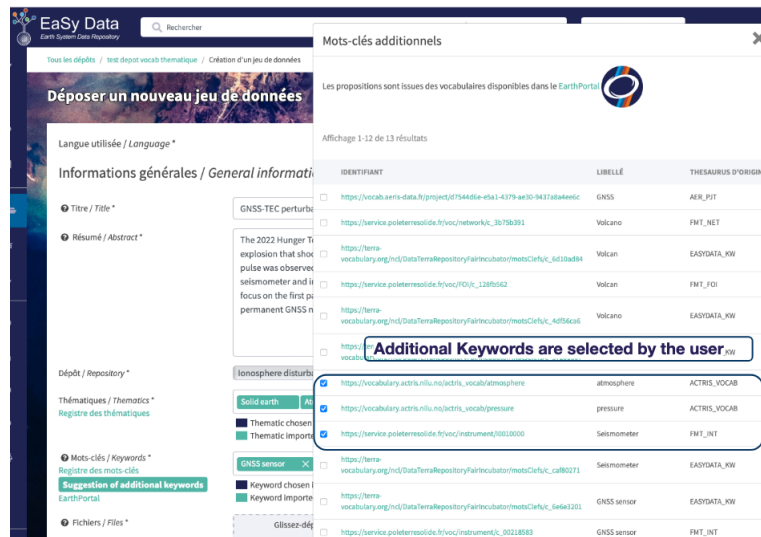


Fig. 10. Step 3 : User selection of additional keywords of interest

Finally, the new keywords are added to the metadata element in a different colour and stored in the corresponding XML file.

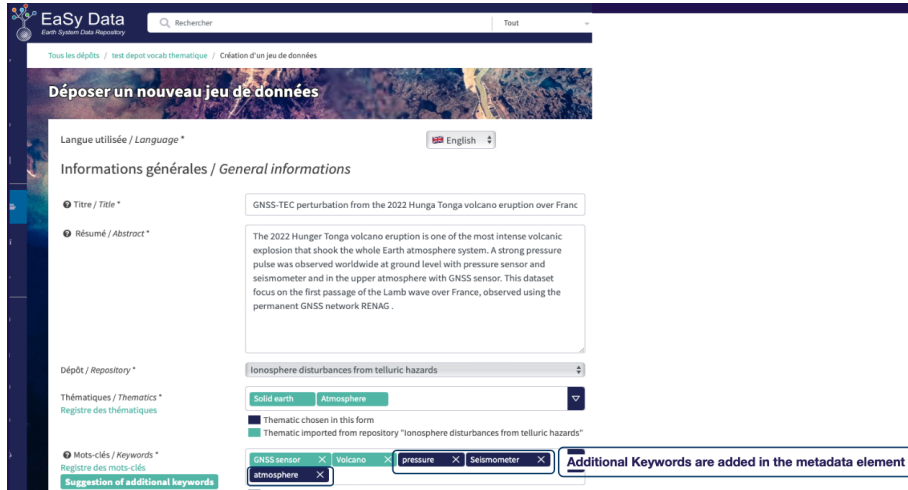


Fig. 11. Step 4 : Keyword inclusion in metadata elements

## 6 Conclusions and perspectives

These results demonstrate that FAIRness, in particular the interoperability criteria, can be significantly improved in a thematic data repository containing spatio-temporal data. The work presented in this paper focused on improving the deposit process by substituting missing relevant metadata entries with the use of controlled vocabularies, in accordance with criterion I1 of the FAIR Principles. Adherence to criterion I2, which requires that the vocabularies themselves need to be FAIR, led to the integration of EaSy Data with EarthPortal, a FAIR catalogue of semantic artefacts, to provide semantic mappings for user-entered terms.

Further improvements to the EaSy Data repository, facilitated by its connection to EarthPortal, are currently under development. These include (i) optimisation of the harvesting process and (ii) enhancements to the search process. As EaSy Data is the national thematic repository for environmental data, it is also intended to host data related to biodiversity and agronomy. It is therefore essential to establish links with semantic artefact catalogues in other domains.

EarthPortal is a member of the ontoportalAlliance [25], which brings together several semantic portals from different domains. Discussions are underway with other semantic artifact catalogues, such as BiodivPortal [26], EcoPortal [27] and AgroPortal [28], to ensure that relevant data is also referenced by domain-specific vocabularies. As part of this alliance, technical discussions are underway to ensure that the connection to these other portals will ultimately be seamless for third-party applications that rely exclusively on the EarthPortal API.

**Acknowledgments.** This work was funded by the EOSC FAIR-IMPACT project (grant number 101057344) and the DATA TERRA Research Infrastructure.

## References

1. Data Terra Homepage, <https://data-terra.org>, last accessed 2024/08/22
2. EaSy Data Repository, <https://easydata.earth>, last accessed 2024/08/22
3. The FAIR Guiding Principles for scientific data management and stewardship, Wilkinson, Mark D and Dumontier, Michel and Aalbersberg, IJsbrand Jan and Appleton, Gabrielle and Axton, Myles and Baak, Arie and Blomberg, Niklas and Boiten, Jan-Willem and da Silva Santos, Luiz Bonino and Bourne, Philip E and others, journal of Scientific data, Nature Publishing Group, volume 3, 2016
4. The Earth-Portal, an ontology repository for the Earth System semantic artefacts, Guillaume Alviset and Christelle Pierkot and Marine Vernet, Onto4FAIR 2023 Workshops, collocated with 19th International Conference on Semantic Systems, SEMANTiCS 2023 <https://earthportal.eu>
5. EOSC FAIR-EASE project, <https://fairease.eu>, last accessed 2024/08/22
6. EOSC Blue Cloud project, <https://blue-cloud.org/>, last accessed 2024/08/22
7. Guide principes FAIR du pôle Odatis, E. Quimbert, M. Fichaut, G. Maudire <https://doi.org/10.13155/87107>
8. Implémenter les principes FAIR dans le système d'information Theia/OZCAR, <https://anr.fr/fr/lanr/engagements/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-fairtois/>, last accessed 2024/08/22
9. Semantics for Fair, usage-oriented semantic enrichment of open data, <https://www.irit.fr/semantics4fair/>, last accessed 2024/08/22
10. Bridge Research through Interoperable Data Governance and Environments, <https://anr.fr/fr/lanr/engagements/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-bridge/>, last accessed 2024/08/22
11. European Open Science Cloud, <https://eosc.eu/>, last accessed 2024/08/22
12. , , last accessed 2024/08/22
13. EOSC FAIR-IMPACT project, <https://fair-impact.eu/>, last accessed 2024/08/22
14. ENVRI-FAIR project, <https://envri.eu/envri-fair/>, last accessed 2024/08/22
15. Reasearch Data Alliance, <https://www.rd-alliance.org/>, last accessed 2024/08/22
16. Une méthodologie et un outil d'évaluation du niveau de FAIRness pour les ressources sémantiques : le cas d'AgroPortal. Emna Amdouni, Clement Jonquet. IC 2021
17. D2.2 FAIR Semantics: First recommendations. Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväläihö, H., Coen, G., & Staiger, C. (1.0). FAIRsFAIR. <https://doi.org/10.5281/zenodo.5361930>
18. Ten simple rules for making a vocabulary FAIR. Coxid, S.J.D., Gonzalez-Beltrani, A.N., Magagna, B., Marinescu, M.-C. 2021. <https://doi.org/10.1371/journal.pcbi.1009041>.
19. Coming to Terms with FAIR Ontologies. M. Poveda-Villalón, P. Espinoza-Arias, D. Garijo, et O. Corcho, in Knowledge Engineering and Knowledge Management, Cham, 2020, p. 255-270. doi:10.1007/978-3-030-61244-3\_18
20. Ontology Repositories and Semantic Artefact Catalogues with the OntoPortal Technology. Clement Jonquet, John Graybeal, Syphax Bouazzouni, Michael Dorf, Nicola Fiore, et al. The Semantic Web - ISWC 2023 [https://doi.org/10.1007/978-3-031-47243-5\\_3](https://doi.org/10.1007/978-3-031-47243-5_3)
21. Eosc interoperability framework. Corcho, O. et al. Publications Office of the European Union, 2021, <https://doi.org/10.2777/620649>.
22. <https://eosc.eu/advisory-groups/fair-metrics-and-data-quality>
23. EPOS research infrastructure <https://www.epos-eu.org/>, last accessed 2024/08/22
24. ACTRIS research infrastructure <https://www.actris.eu/>
25. <https://ontoportals.org/>
26. <https://biodivportal.gfbio.dev/>
27. <https://ecoportal.lifewatch.eu/>
28. <https://agroportal.lirmm.fr/>