



HAL
open science

Améliorer l'interopérabilité des entrepôts de données avec l'utilisation de catalogues d'artefacts sémantiques

Christelle Pierkot, Guillaume Alviset, Hélène Bressan

► To cite this version:

Christelle Pierkot, Guillaume Alviset, Hélène Bressan. Améliorer l'interopérabilité des entrepôts de données avec l'utilisation de catalogues d'artefacts sémantiques. EGC 2025, Jan 2025, Strasbourg, France. hal-04908697

HAL Id: hal-04908697

<https://hal.science/hal-04908697v1>

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 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.



Distributed under a Creative Commons Attribution 4.0 International License

Améliorer interopérabilité des entrepôts de données avec l'utilisation de catalogues d'artefacts sémantiques

Résumé. l'étude du système Terre regroupe plusieurs disciplines dont les collaborations se multiplient pour répondre à des problématiques de plus en plus complexes, comme celles relatives au changement climatique. Data Terra, l'infrastructure de recherche nationale en charge de la gestion des données du Système Terre offre des services de découverte, d'accès et de diffusion des données, permettant aux chercheurs de mener à bien leurs travaux scientifiques. Parmi ces services, l'entrepôt de données national EaSy Data, facilite le partage et la valorisation des données de la recherche, et l'EarthPortal, un catalogue FAIR d'artefacts sémantiques, encourage l'utilisation et la diffusion de vocabulaires contrôlés et d'ontologies.

Malgré le fait que les gestionnaires de données sont convaincus du bien fondé des principes FAIR et des apports de leur application, leurs mises en œuvre restent difficiles à atteindre. C'est le cas en particulier pour le critère d'interopérabilité alors qu'il est un enjeu majeur, essentiel pour garantir que les données puissent être efficacement combinées, interprétées et utilisées à la fois par les humains et les machines. Pour satisfaire aux critères d'interopérabilité des principes FAIR, il est impératif que les données et les métadonnées utilisent des vocabulaires contrôlés et des modèles de représentation structurés.

Dans cet article, nous montrons comment l'utilisation d'un catalogue d'artefacts sémantiques peut considérablement renforcer l'interopérabilité d'un entrepôt de données. Nous illustrons cette démarche par l'intégration de l'entrepôt de données EaSy Data avec le catalogue sémantique EarthPortal, ce qui permet d'enrichir les métadonnées par des annotations sémantiques tout en améliorant l'expérience utilisateur.

1 Introduction

l'étude du système Terre regroupe plusieurs disciplines dont les collaborations se multiplient pour répondre à des problématiques de plus en plus complexes, comme celles relatives au changement climatique. Data Terra ¹ est l'infrastructure de recherche en charge de la gestion des données de 5 compartiments majeurs du Système Terre : Océan, Atmosphère, Terre Solide, Surface continentale et Biodiversité. En plus de services distribués d'infrastructure, Data Terra offre également

1. Data Terra : <https://data-terra.org>

des services de découverte, d'accès et de diffusion aux données, permettant aux chercheurs de mener à bien leurs travaux scientifiques. Parmi ces services, l'entrepôt de données national EaSy Data, permet aux chercheurs de partager leurs données en vue de diffusion et de valorisation auprès d'autres utilisateurs potentiels. Les principes FAIR définis par Wilkinson et al. (2016) pour permettre une meilleure gestion des données dans les entrepôts de données, à travers 4 principes : *Findability*, *Accessibility*, *Interoperability* and *Reusability*, sont reconnus d'utilité par les gestionnaires des données du Système Terre et plusieurs initiatives et projets ont été menés pour améliorer ceux-ci (EOSC FAIR-EASE², Blue-Cloud³, Copilote⁴). En particulier, le respect du principe d'Interopérabilité dans les entrepôts de données garantit que les données sont téléchargeables, utilisables, compréhensibles, et combinables avec d'autres données, à la fois par des humains et des machines. Cependant, les retours d'expérience montrent que le principe le plus difficile à atteindre quant il s'agit de partager et réutiliser les données est celui d'Interopérabilité. Les critères I1 et I2 des principes FAIR, spécifient que : (i) la description des données doit suivre un vocabulaire contrôlé et un modèle de représentation servant à le structurer et (ii) les données et les métadonnées utilisent des vocabulaires qui respectent eux-mêmes les principes FAIR. Ces vocabulaires, documentés et accessibles via un identifiant permanent (PID), doivent être lisibles par machine. Par conséquent, l'utilisation avisée de la sémantique dans les entrepôts de données est essentielle pour renforcer leur interopérabilité. Dans cet article, nous montrons, à travers un exemple d'application mis en oeuvre dans l'entrepôt de données EaSy Data⁵, comment l'utilisation d'EarthPortal⁶, un catalogue/référentiel d'artefacts sémantiques, peut non seulement améliorer l'interopérabilité des entrepôts de données, mais aussi simplifier l'expérience de dépôt pour les utilisateurs. Le papier est structuré de la manière suivante : Tout d'abord, sur la base des recommandations des principes FAIR, nous analysons comment la sémantique pourrait être utilisée pour améliorer l'interopérabilité des entrepôts de données. Ensuite, nous présentons EaSy Data, un entrepôt de données thématiques, et EarthPortal, un catalogue d'artefacts sémantiques, et décrivons les différentes utilisations possibles de l'intégration des deux outils. Nous présentons ensuite l'exemple de la connexion entre EarthPortal et EaSy Data et la valeur ajoutée qu'elle apporte. Enfin, nous proposons quelques perspectives d'avenir pour ce travail.

2 Interopérabilité des Entrepôts de Données : l'impact des Principes FAIR et des technologies Sémantiques

Les principes FAIR ont joué un rôle crucial dans l'établissement des meilleures pratiques de gestion des données au sein des communautés scientifiques, obtenant un soutien important de la part des organismes de financement de la recherche pour encourager leur adoption. En France, le programme de recherche Science Ouverte de l'ANR a soutenu divers projets centrés sur ces principes, tels que FairTOIS⁷, Semantics4Fair⁸, BRIDGE⁹, et Copilote. Le programme Euro-

2. EOSC FAIR-EASE project, <https://fairease.eu>

3. EOSC Blue Cloud project, <https://blue-cloud.org/>

4. Copilote : <https://anr.fr/fr/lanr/engagements/la-science-ouverte/les-projets-laureats-de-lappel-flash-science-ouverte/projet-copilote/>

5. EaSy Data : Earth System Data Repository <https://easydata.earth>

6. EarthPortal : <https://earthportal.eu>

7. FairTOIS : 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/>

8. Semantics4FAIR : Semantics for Fair, usage-oriented semantic enrichment of open data, <https://www.irit.fr/semantics4fair/>

9. BRIDGE : 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/>

pean Open Science Cloud¹⁰ est également fondamental dans la mise en œuvre de ces principes, soutenant des projets majeurs tels que FAIRsFAIR¹¹, FAIR-IMPACT¹², ENVRI-FAIR¹³, et FAIR-EASE, ainsi qu'en constituant des groupes de travail sur les données FAIR, tels que la Task Force sur les métriques FAIR et la qualité des données¹⁴ (2021-2023) et plus récemment, la Task Force sur les objets numériques et les métriques FAIR (2024-). À l'échelle internationale, la Research Data Alliance¹⁵ se distingue par son engagement continu envers les initiatives liées aux principes FAIR. Plusieurs groupes de travail ont émergé ces dernières années (tels que FAIR4RS, FAIR4ML, FAIR4VRE et FAIR mappings), témoignant de l'intérêt croissant de la communauté scientifique, y compris dans les sciences de la Terre, à intégrer ces principes dans leurs domaines respectifs.

Le respect des principes FAIR dans les entrepôts de données garantit que les données sont Faciles à trouver, Accessibles, Interopérables, Réutilisables, et qu'elles peuvent être intégrées avec d'autres ensembles de données. Les critères d'interopérabilité des principes FAIR, qui stipulent que (I1) la description des données doit suivre un vocabulaire contrôlé et un modèle de représentation pour structurer ce vocabulaire, et (I2) les données et métadonnées doivent utiliser des vocabulaires conformes aux principes FAIR, sont essentiels pour la mise en œuvre des entrepôts de données, garantissant que les données soient compréhensibles et utilisables par les humains et les machines. Cependant, comme le souligne Amdouni et al. (2021), les principes sous-jacents à l'Interopérabilité sont souvent considérés comme les plus difficiles à atteindre et les plus cruciaux pour répondre aux exigences FAIR. Pour répondre au critère I1, les métadonnées qui permettent de décrire explicitement les données dans les entrepôts doivent suivre un schéma standardisé, commun, partagé et sémantiquement riche. Dans les entrepôts de données, les artefacts sémantiques (un terme générique proposé par le projet FAIRsFAIR pour désigner les objets sémantiques tels que les vocabulaires, ontologies, etc. LeFranc et al. (2020) sont devenus essentiels pour respecter les critères d'interopérabilité des principes FAIR.

Dans le domaine des sciences de la Terre, divers artefacts sémantiques (AS), tels que des vocabulaires et des ontologies, sont utilisés pour représenter et annoter les données de manière standardisée. Par exemple, l'ontologie SWEET (Semantic Web for Earth and Environment Technology Ontology) et les thésaurus du GCMD (Global Change Master Directory) comptent parmi les références majeures. Cependant, ces artefacts sémantiques de haut niveau ne couvrent pas de manière exhaustive toutes les caractéristiques du système terre. Par conséquent, des initiatives ont été mises en place pour développer des artefacts sémantiques plus spécifiquement adaptés. L'infrastructure de recherche Data Terra a spécifié des vocabulaires afin qu'ils puissent être utilisés dans les catalogues des pôles de données. Le thésaurus EaSy Data, un vocabulaire unifié, a été créé pour être utilisé dans l'entrepôt national de données thématiques EaSy Data, qui héberge les données de recherche de longue traîne des composantes du système Terre. À l'échelle européenne, l'infrastructure Actris, dédiée à la recherche sur le climat et la qualité de l'air, a développé un vocabulaire contrôlé qu'elle utilise dans son catalogue. De même, EPOS, l'infrastructure pour la composante Terre solide, utilise un vocabulaire dans les services d'EPOS-MSL¹⁶.

Pour faire face à la multiplication des artefacts sémantiques et respecter le critère I2, Coxid et al. (2021), Poveda-Villalón et al. (2020) proposent dix règles pour rendre les artefacts sémantiques eux-mêmes FAIR (par exemple, assigner un identifiant unique et persistant, ajouter des

projet-bridge/

10. EOSC : European Open Science Cloud, <https://eosc.eu/>

11. FAIRsFAIR

12. FAIR-IMPACT : EOSC FAIR-IMPACT project, <https://fair-impact.eu/>

13. ENVRI-FAIR : ENVRI-FAIR project, <https://envri.eu/envri-fair/>

14. FAIR metrics Task Force : <https://eosc.eu/advisory-groups/fair-metrics-and-data-quality>

15. RDA : Research Data Alliance, <https://www.rd-alliance.org/>

16. EPOS Thematic Core Services Multi-Scale Laboratories

métadonnées, enregistrer l'artefact sémantique dans un référentiel sémantique, etc.). Jonquet et al. (2023) recommandent l'utilisation de catalogues permettant le partage, la diffusion et la réutilisation des artefacts sémantiques. Sur la base de ces recommandations, nous avons traité un exemple concret pour montrer comment un catalogue d'artefacts sémantiques peut améliorer l'interopérabilité des référentiels de données spatiales.

3 Gestion de l'interopérabilité dans le référentiel EaSy Data

EaSy Data est l'entrepôt de données national français de confiance dédié aux données de recherche liées aux sciences de la Terre et de l'environnement.

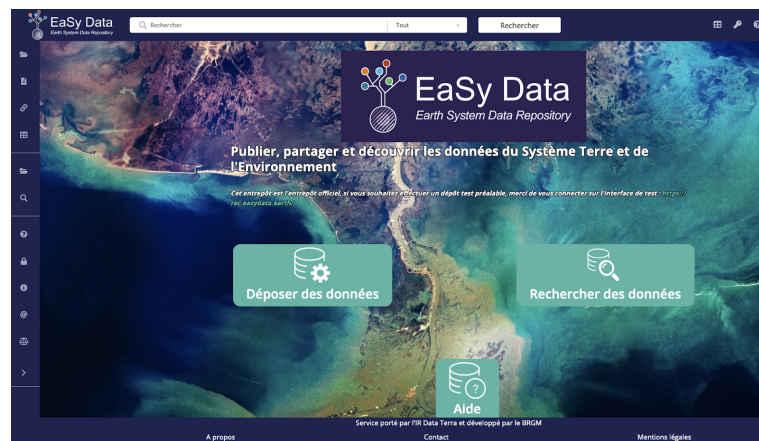


FIG. 1 – Interface principale d'EaSy Data.

Grâce à EaSy Data, un chercheur peut déposer des données provenant de tous les domaines des sciences de la Terre et demander un DOI. L'approbation est soumise à un processus de modération réalisé par des experts de différentes disciplines. Une interface de recherche est également disponible, permettant aux utilisateurs de retrouver facilement les données.

Conformément aux standards internationaux, EaSy Data repose sur un schéma de métadonnées standardisé, en utilisant la norme ISO 19115 pour la description des données. La gestion des métadonnées est assurée via l'outil Geonetwork. De plus, une couche d'application web a été développée pour faciliter le dépôt des données ainsi que la gestion des fonctionnalités associées (comme la recherche et la gestion des droits).

Pour se conformer aux principes FAIR, en particulier ceux liés à l'interopérabilité, plusieurs éléments de métadonnées dans EaSy Data intègrent des artefacts sémantiques. Ces artefacts sont exposés selon différents formats du web sémantique et sont accessibles via des API. Alors que certains artefacts sémantiques proviennent de sources publiques accessibles en ligne, d'autres ont été créés spécifiquement pour l'entrepôt EaSy Data. Le tableau ci-dessous présente les éléments de métadonnées qui, dès la première version d'EaSy Data, ont été initialement associés à un ou plusieurs artefacts sémantiques.

Les vocabulaires utilisés pour les éléments de métadonnées "Thématiques" et "Mots-clés" ont été définis par des scientifiques. Cependant, ces vocabulaires restent incomplets et ne reflètent pas entièrement la diversité des données déposées dans EaSy Data. d'autres vocabulaires existent et pourraient être utilisés pour compléter ces vocabulaires, tels que certains thésaurus produits

Élément de métadonnées	Description de l'élément de métadonnées	Exemple	Artefact Sémantique utilisé
Thématiques	Thématiques générales ou domaines auxquels les données appartiennent	sismologie, atmosphère	EaSyData_Thématique
Mots-clés	Mots-clés décrivant les caractéristiques spécifiques du jeu de données.	Conductivité de l'eau de surface, signal sismique	EaSyData_Keywords
Organisation	Organisations de recherche	CNRS, IRD	ROR
Licence	Licence de distribution des données	CC-BY-4.0, etalab-2.0	SPDX Licenses
Description géographique	Description géographique de l'étendue spatiale du jeu de données	France, Californie	Geonames
Codes	Listes de codes pour la description des jeux de métadonnées conformes à la norme ISO 19115	Point de contact, Responsable (Role Code) - vecteur, grille (SpatialRepresentation-TypeCode)	ISO 19115 Codelists

TAB. 1 – Éléments de métadonnées liés à un artefact sémantique dans EaSy Data

par l'infrastructure nationale Data Terra ou par des infrastructures européennes comme ceux définis par EPOS¹⁷ ou ACTRIS¹⁸.

4 Améliorer l'interopérabilité grâce à EarthPortal

EarthPortal [Pierkot et al. (2023)], est un catalogue d'artefacts sémantiques thématiques et un référentiel sémantique dédié aux sciences de la Terre, basé sur la technologie OntoPortal.

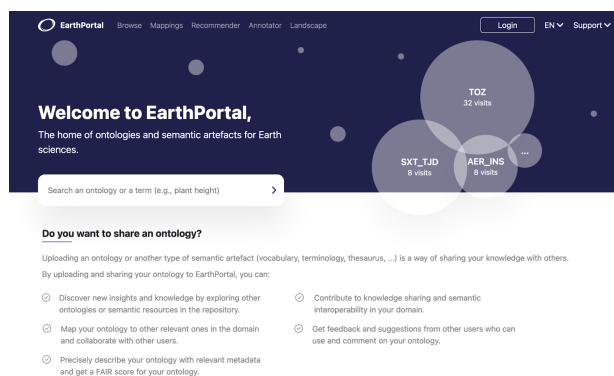


FIG. 2 – Interface principale d'EarthPortal.

17. EPOS research infrastructure <https://www.epos-eu.org/>

18. ACTRIS research infrastructure <https://www.actris.eu/>

Améliorer l'Interopérabilité des Données grâce à un Catalogue d'Artefacts Sémantiques

Développé dans le cadre du projet EOSC FAIR-IMPACT, sa mise en œuvre suit les recommandations nationales et européennes, avec pour objectif d'améliorer la diffusion et la réutilisation des artefacts sémantiques dans un contexte de recherche interdisciplinaire. EarthPortal assure également la traçabilité et la reconnaissance de la paternité des vocabulaires, offrant ainsi une meilleure reconnaissance à leurs créateurs. Au niveau européen, les recommandations de la Task Force "Interopérabilité Sémantique" mettent en évidence l'importance de reconnaître les catalogues d'artefacts sémantiques (tels que décrits dans le cadre d'interopérabilité de l'European Open Science Cloud - EOSC) comme un élément crucial pour la durabilité à long terme de toute infrastructure de données de recherche [Corcho et al. (2021)].

EarthPortal héberge des artefacts sémantiques relatifs aux sciences de la Terre et de l'environnement, et peut être utilisé par des applications externes via son API REST. Ces artefacts incluent des vocabulaires contrôlés au format SKOS et des ontologies spécifiées en OWL. Ils sont organisés en catégories et groupes dans le catalogue. Des filtres peuvent être appliqués à ces éléments selon les besoins. Les thésaurus de Data Terra, ainsi que les vocabulaires produits par des infrastructures européennes et d'autres artefacts sémantiques plus génériques mais largement utilisés (comme SOSA, SWEET, etc.), sont actuellement disponibles dans EarthPortal.

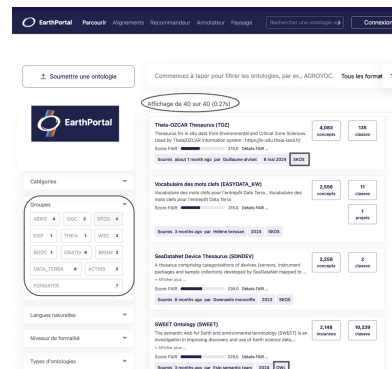


FIG. 3 – Artefacts sémantiques dans EarthPortal.

En plus de faciliter la diffusion et la réutilisation des artefacts sémantiques des sciences de la Terre, EarthPortal propose des outils exploitables par des applications tierces : l'Annotateur, qui suggère des termes basés sur du texte ou des mots-clés ; le Recommandeur, qui propose des artefacts sémantiques pertinents selon le texte ou les mots-clés saisis ; et l'outil Mappings, qui génère, stocke et affiche les correspondances entre artefacts sémantiques.

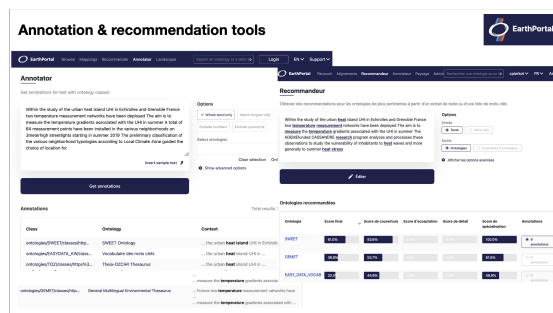


FIG. 4 – Services d'annotation et de recommandation proposés par EarthPortal.

L'Annotateur est particulièrement intéressant, car il permet d'identifier des concepts au sein des artefacts sémantiques hébergés sur EarthPortal à partir de mots-clés ou de texte.

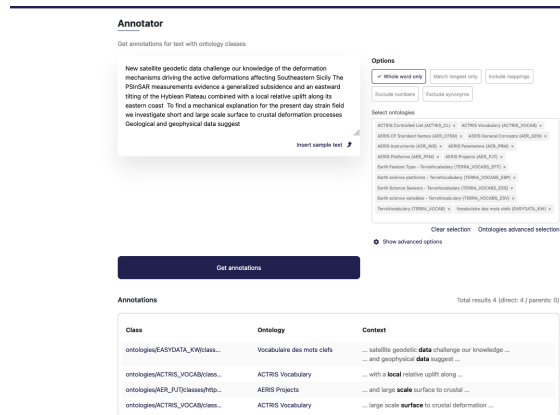


FIG. 5 – *Outil Annotator.*

Des filtres peuvent être appliqués pour limiter la sélection aux artefacts pertinents. Il est notamment possible de filtrer par catégorie pour obtenir des résultats spécifiques à un sujet donné. Dans l'exemple suivant, nous choisissons de limiter la sélection aux artefacts sémantiques liés au thème de l'Atmosphère.

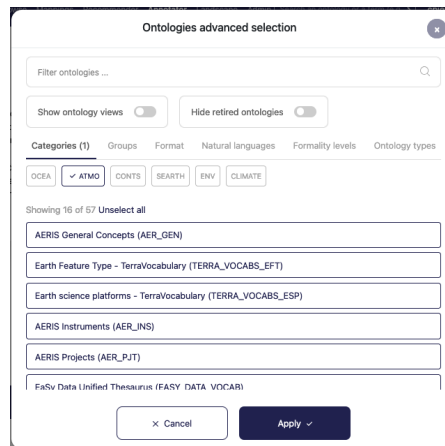


FIG. 6 – *Filtrage des artefacts sémantiques pertinents par catégories.*

Grâce à son API, des applications tierces comme EaSy Data peuvent se connecter à EarthPortal pour utiliser les artefacts sémantiques disponibles et exploiter les outils proposés.

5 Utilisation d'EarthPortal dans EaSy Data : une application concrète

Nous avons identifié trois cas d'usage potentiels pour l'intégration d'EarthPortal au sein d'EaSy Data :

— *Dépôt de données :*

Nous proposons d'améliorer le processus de saisie des métadonnées, en particulier en ce qui concerne les mots-clés, grâce à l'utilisation d'artefacts sémantiques disponibles dans EarthPortal. Actuellement, dans EaSy Data, les mots-clés sont définis par des experts thématiques sans hiérarchie structurée ni classification précise des domaines. Par conséquent, ces mots-clés sont souvent incomplets ou inappropriés pour certains jeux de données. Par exemple, les mots-clés liés aux études atmosphériques peuvent ne pas convenir aux jeux de données concernant les glissements de terrain. Pour pallier cette limitation, nous proposons de connecter EaSy Data à EarthPortal afin d'offrir aux chercheurs des mots-clés pertinents et adaptés au contexte lors du dépôt de leurs jeux de données.

— *Recherche de jeux de données :*

Les filtres de recherche dans EaSy Data sont actuellement limités par leur dépendance aux vocabulaires spécifiques au référentiel, restreignant ainsi l'étendue des résultats. Le web sémantique, grâce aux correspondances entre concepts, permet d'établir des relations entre entités à travers différents artefacts sémantiques. La fonctionnalité de "mapping" d'EarthPortal identifie automatiquement des correspondances entre les vocabulaires de son catalogue. Nous envisageons de tirer parti de cette capacité pour améliorer la recherche à facettes dans EaSy Data, élargissant ainsi la portée et la pertinence des résultats de recherche.

— *Moissonnage des métadonnées :*

— EaSy Data a pour mission de moissonner d'autres entrepôts de données, tels que Seanoë, exclusivement dédié aux données marines, pour rendre ces informations accessibles à une échelle pluridisciplinaire. Parfois, ces entrepôts n'utilisent pas de vocabulaires contrôlés, et proposent uniquement du texte libre pour définir certains éléments de métadonnées comme les mots clés, ce qui est contraire aux spécifications d'EaSy Data. Pour résoudre ce problème, nous proposons d'utiliser EarthPortal pour identifier les correspondances sémantiques pour ces éléments de métadonnées moissonnés, enrichissant ainsi leur qualité sémantique au sein d'EaSy Data.

— Par ailleurs, EaSy Data sera lui-même moissonné par d'autres catalogues, tels que l'infrastructure européenne EPOS, qui vise à intégrer les jeux de données d'EaSy Data dans un contexte plus large. Cependant, ces catalogues ne partagent pas le même vocabulaire qu'EaSy Data, ce qui limite les possibilités d'intégration. Pour faciliter l'exploitation des métadonnées d'EaSy Data dans ces catalogues externes, il est essentiel d'établir des correspondances entre les vocabulaires respectifs. En utilisant EarthPortal pour stocker les vocabulaires des différents référentiels et en exploitant ses fonctionnalités de mapping et d'annotation, nous pouvons significativement réduire le risque de perte d'information.

Dans la section suivante, nous nous concentrons exclusivement sur les résultats du premier cas d'usage : le dépôt de données.

5.1 Amélioration de l'expérience de dépôt de données

Lorsque l'utilisateur souhaite soumettre un jeu de données dans EaSy Data, il/elle doit d'abord fournir un titre et un résumé, puis sélectionner des thématiques et des mots-clés pertinents à partir du thésaurus d'EaSy Data. Si les mots-clés disponibles ne sont pas suffisamment pertinents, l'utilisateur peut choisir de les enrichir en utilisant un bouton désigné dans l'interface.

Cette action déclenche une requête vers l'API d'EarthPortal. Les artefacts sémantiques cibles sont filtrés en fonction des thématiques sélectionnées par l'utilisateur dans EaSy Data, réduisant ainsi le risque de mauvaise interprétation des termes hors du domaine souhaité.

The screenshot shows the EaSy Data web interface. At the top, there is a search bar and a language selector set to English. The main heading is "Déposer un nouveau jeu de données". Below this, there are several sections:

- Langue utilisée / Language ***: English
- Informations générales / General informations**:
 - Titre / Title ***: GNSS-TEC perturbation from the 2022 Hunga Tonga volcano eruption over France
 - Résumé / Abstract ***: The 2022 Hunga Tonga volcano eruption is one of the most intense volcanic explosion that shook the whole Earth atmosphere system. A strong pressure pulse was observed worldwide at ground level with pressure sensor and seismometer and in the upper atmosphere with GNSS sensor. This dataset focus on the first passage of the Lamb wave over France, observed using the permanent GNSS network RENAG.
 - Dépôt / Repository ***: Ionosphere disturbances from telluric hazards
 - Thématiques / Thematics ***: Solid earth, Atmosphere
 - Mots-clés / Keywords ***: GNSS sensor, Volcano
- Suggestion of additional keywords**: A button with a red border and a red background, highlighted by a red box.

FIG. 7 – Étape 1 : Clic sur le bouton de suggestion de mots-clés

La requête associée est la suivante :

$https://data.earthportal.eu/categories/{Topic}?display=ontologies$ où Topic est la thématique sélectionnée par l'utilisateur dans le formulaire d'EaSy Data.

EarthPortal traite ensuite ce résumé via son service d'annotation pour proposer des concepts sémantiques. La requête associée est la suivante :

$https://data.earthportal.eu/annotator?text={Abstract}\&ontologies={List\ of\ SA}$ où :

- Abstract est le résumé fourni par l'utilisateur dans le formulaire d'EaSy Data
- List of SA est le résultat de la requête précédente

Le résultat de cette requête est un fichier JSON contenant toutes les correspondances entre les termes donnés dans le résumé d'EaSy Data et les concepts trouvés dans les artefacts sémantiques hébergés dans EarthPortal, filtrés par catégorie, comme montré dans l'exemple ci-dessous :

```

https://www.easydata.earth/back/public-api/v1/earthPortal/annotatorSearch?
JSON  Données brutes  En-têtes
Enregistrer Copier Tout réduire Tout développer Filter le JSON
> 0: (-)
> 1: (-)
> 2: (-)
> 3: (-)
> 4: (-)
- 5:
  - id: "https://vocabulary.actris.nslu.nu/actris_vocab/atmosphere"
  label: "atmosphere"
  ontologyAcronym: "ACTRIS_VOCAB"
- 6:
  - id: "https://vocabulary.actris.nslu.nu/actris_vocab/pressure"
  label: "pressure"
  ontologyAcronym: "ACTRIS_VOCAB"
> 7: (-)
- 8:
  - id: "https://service.unlterresolide.fr/voc/instrument/1001000"
  label: "Seismometer"
  ontologyAcronym: "FHT_INT"
> 9: (-)
> 10: (-)
> 11: (-)
> 12: (-)
> 13: (-)
- 14:
  - id: "https://terra-vocabulary.org/incl/DataTerraRepository/incubator/motsClefs/c_1127290"
  label: "wave"
  ontologyAcronym: "EASYSYDATA_VOC"

```

FIG. 8 – Extrait du fichier JSON résultant de l'annotation de l'exemple ci-dessus

Améliorer l'Interopérabilité des Données grâce à un Catalogue d'Artefacts Sémantiques

EaSy Data interprète ce fichier JSON et propose des mots-clés supplémentaires dans une nouvelle fenêtre, permettant à l'utilisateur de sélectionner ceux qui l'intéressent.

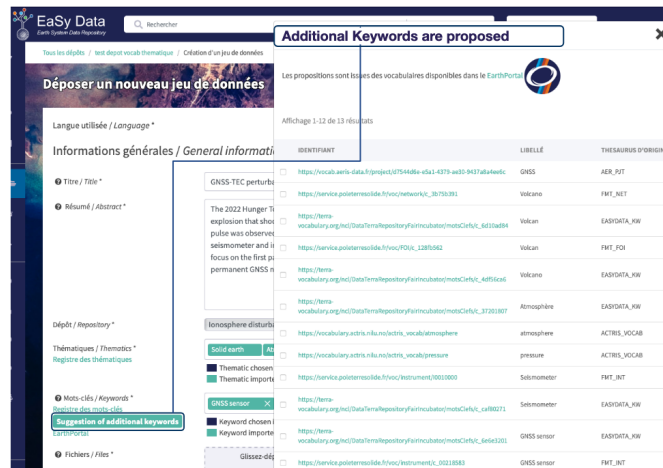


FIG. 9 – Étape 2 : Suggestion de mots-clés supplémentaires basée sur le résultat de l'annotateur EarthPortal

L'utilisateur sélectionne un ou plusieurs mots-clés supplémentaires et envoie la demande à EaSy Data.

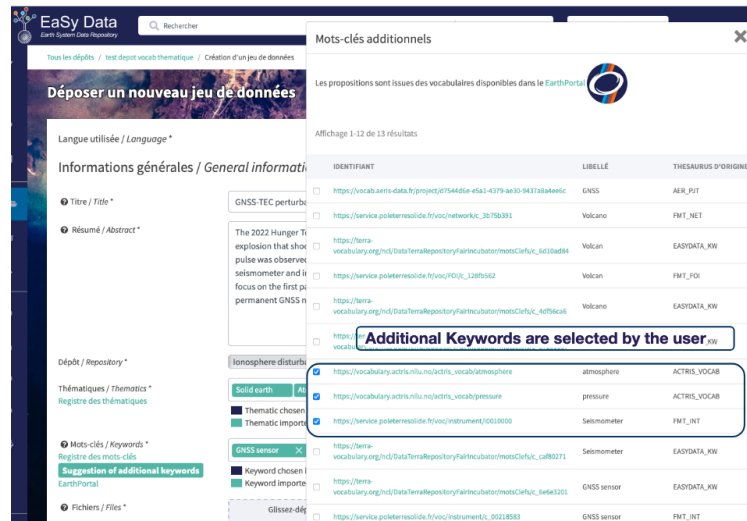


FIG. 10 – Étape 3 : Sélection par l'utilisateur des mots-clés supplémentaires d'intérêt

Enfin, les nouveaux mots-clés sont ajoutés à l'élément de métadonnées dans une couleur différente et stockés dans le fichier XML correspondant.

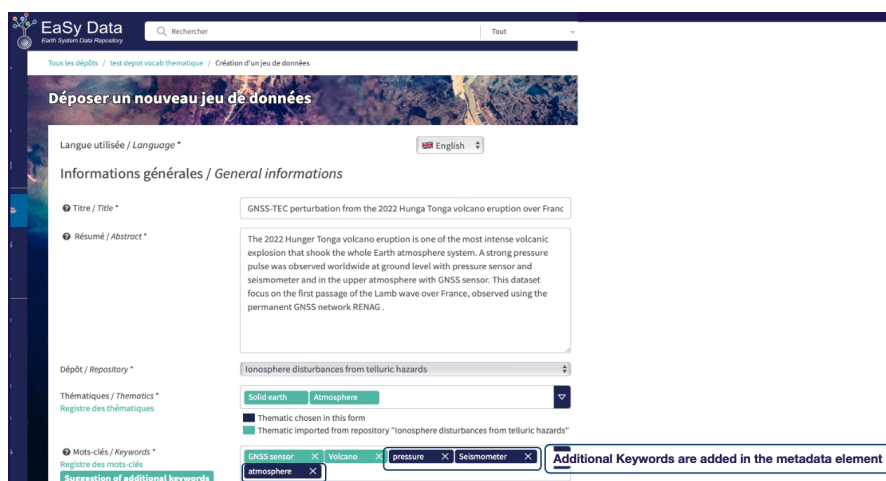


FIG. 11 – *Étape 4 : Inclusion des mots-clés dans les éléments de métadonnées*

6 Conclusions et perspectives

Les résultats présentés démontrent que la conformité aux principes FAIR, en particulier les critères d'interopérabilité, peut être considérablement améliorée dans un entrepôt de données thématiques contenant des données spatio-temporelles. Les travaux décrits dans cet article se sont concentrés sur l'amélioration du processus de dépôt de données en remplaçant les entrées de métadonnées manquantes ou inadéquates par l'utilisation de vocabulaires contrôlés, conformément au critère I1 des principes FAIR. L'adhésion au critère I2, qui stipule que les vocabulaires eux-mêmes doivent être FAIR, a conduit à l'intégration d'EaSy Data avec EarthPortal, un catalogue FAIR d'artefacts sémantiques, afin de fournir des correspondances sémantiques pour les termes saisis par les utilisateurs. Des améliorations supplémentaires du référentiel EaSy Data, rendues possibles grâce à sa connexion à EarthPortal, sont actuellement en cours de développement. Elles incluent (i) l'optimisation du processus de moissonnage et (ii) des améliorations du processus de recherche. En tant qu'entrepôt thématique national pour les données environnementales, EaSy Data a également vocation à héberger des données relatives à la biodiversité et à l'agronomie. Il est donc essentiel d'établir des liens avec des catalogues d'artefacts sémantiques dans d'autres domaines. EarthPortal fait partie de l'ontoportail Alliance¹⁹, qui regroupe plusieurs portails sémantiques issus de différents domaines. Des discussions sont en cours avec d'autres catalogues d'artefacts sémantiques, tels que BiodivPortal²⁰, EcoPortal²¹ et AgroPortal²², afin de garantir que les données pertinentes soient également référencées par des vocabulaires spécifiques à ces domaines. Dans le cadre de cette alliance, des discussions techniques sont en cours pour garantir que la connexion à ces autres portails sera finalement transparente pour les applications tierces qui s'appuient exclusivement sur l'API d'EarthPortal.

19. OntoPortalAlliance : <https://ontoportail.org/>

20. BiodivPortal : <https://biodivportal.gfbio.dev/>

21. EcoPortal : <https://ecoportal.lifewatch.eu/>

22. AgroPortal : <https://agroportal.lirmm.fr/>

Références

- 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.
- The Earth-Portal, an ontology repository for the Earth System semantic artefacts, Christelle Pierkot and Guillaume Alviset and Marine Vernet, *Onto4FAIR 2023 Workshops, collocated with 19th International Conference on Semantic Systems, SEMANTiCS 2023* <https://earthportal.eu>
- Guide principes FAIR du pôle Odatis, E. Quimbert, M. Fichaut, G. Maudire <https://doi.org/10.13155/87107>
- 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
- D2.2 FAIR Semantics : First recommendations. Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväsliho, H., Coen, G., & Staiger, C. (1.0). FAIRsFAIR. <https://doi.org/10.5281/zenodo.5361930>
- Ten simple rules for making a vocabulary FAIR. Coxid, S.J.D., Gonzalez-Beltrán, A.N., Magagna, B., Marinescu, M.-C. 2021. <https://doi.org/10.1371/journal.pcbi.1009041>.
- 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
- 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
- Eosc interoperability framework. Corcho, O. et al. Publications Office of the European Union, 2021, <https://doi.org/10.2777/620649>.

Summary

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.