



HAL
open science

De TermSciences à Loterre : comment l’Inist-CNRS a rendu les terminologies ouvertes plus conformes aux principes FAIR

Majid Khayari, Véronique Reszetko, Dominique Vachez, Nathalie Vedovotto, Jérémy Yon, Sophie Aubin

► To cite this version:

Majid Khayari, Véronique Reszetko, Dominique Vachez, Nathalie Vedovotto, Jérémy Yon, et al.. De TermSciences à Loterre : comment l’Inist-CNRS a rendu les terminologies ouvertes plus conformes aux principes FAIR. 2021, 10.15454/djh8-yp20 . hal-03176063

HAL Id: hal-03176063

<https://hal.inrae.fr/hal-03176063v1>

Submitted on 6 Apr 2021

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

De TermSciences à Loterre : comment l'Inist-CNRS a rendu les terminologies ouvertes plus conformes aux principes FAIR

Cet article a été publié le 1^{er} avril 2021 sur le site du projet FooSIN : <https://foosin.fr/de-termsciences-a-loterre-comment-linist-cnrs-a-rendu-les-terminologies-ouvertes-plus-conformes-aux-principes-fair/>

Les membres du projet FooSIN ont interviewé l'Inist-CNRS à propos du passage de **TermSciences à Loterre, deux portails publics exposant des terminologies d'intérêt notamment pour les acteurs de l'agriculture, de l'alimentation et de l'environnement**. En effet, une dizaine de ces terminologies ont été publiées par INRAE. Les objectifs de cet entretien est de montrer en quoi Loterre a progressé par rapport à TermSciences et quel est l'impact de ce changement sur les ressource hébergées en termes de satisfaction des principes FAIR. Suite aux échanges et en s'appuyant sur l'utilisation de la grille d'évaluation SHARC, le projet FooSIN suggère quelques pistes pour aller plus loin dans la démarche. Enfin, vous trouverez dans cet article des pointeurs vers les outils, standards et ressources utilisées pour la FAIRification.

Contenu

Préambule	2
Démarche de FAIRification	2
TermSciences.....	2
Loterre	2
Discussion	3
Evaluation FAIR à l'aide de la grille SHARC.....	5
Références par type et dans l'ordre d'apparition dans le texte.....	8
Lectures et documents de référence	8
Portails et sites web	8
Standards.....	8
Outils et services	9
Terminologies d'intérêt pour l'agriculture et l'alimentation publiées dans Loterre	9

Préambule

Démarche de FAIRification

Entre la création des deux portails TermSciences et Loterre, [les principes FAIR](#) ont été publiés par [Wilkinson en 2016](#). **L'objectif des principes FAIR est de favoriser la découverte, l'accès, l'interopérabilité et la réutilisation des objets numériques.** Leur mise en œuvre au sein d'un système d'information ou d'une ressource implique le choix de certaines technologies, de modèles de représentation standards pour les données et les métadonnées, ou encore de modes de gouvernance des ressources. Ces choix sont souvent appelés « démarche de FAIRification » (ou « démarche de FAIRisation »). **Les terminologies et autres ressources sémantiques ont une place particulière dans la démarche de FAIRification puisqu'elles en sont à la fois un moyen et l'objet.** En effet, le Principe I2 énonce que **“les données et les métadonnées utilisent des vocabulaires qui respectent les principes FAIR”**.

TermSciences

L'Inist-CNRS (Institut de l'Information Scientifique et Technique) a développé en 2005 le **portail terminologique TermSciences** en association avec le LORIA (Laboratoire lorrain de Recherche en Informatique et ses Applications) et l'ATILF (Analyse et Traitement Informatique de la Langue Française). **Le portail avait pour but de valoriser et de mutualiser les ressources terminologiques (lexiques, dictionnaires, thésaurus) des organismes publics de recherche et d'enseignement supérieur pour aboutir à la constitution d'un référentiel terminologique commun.** L'Inist en tant que producteur et utilisateur de terminologies scientifiques y a joué un rôle de coordination. Elle a ainsi fédéré un réseau de partenaires qui ont apporté leur contribution à la constitution et à la gestion du contenu terminologique, et d'experts qui ont mis leurs compétences, leurs expériences et leur savoir-faire au service de cette réalisation. Aujourd'hui, le portail TermSciences est toujours [accessible en ligne](#) à la consultation mais n'est plus alimenté.

Loterre

En 2018, l'Inist-CNRS crée **Loterre** (Linked open terminology resources) pour **permettre l'exposition et le partage de terminologies scientifiques multidisciplinaires et multilingues dans une vision orientée vers les standards du web des données ouvertes et liées (LOD) et les principes FAIR.** Loterre continue à s'enrichir depuis sa création : en mars 2021, 42 ressources terminologiques y sont exposées. L'accès aux ressources, en consultation, interrogation et téléchargement est ouvert à tous. Loterre s'intègre dans le monde des données ouvertes et liées, les terminologies disponibles sur le portail sont donc exploitables directement par des machines via un SPARQL Endpoint en plus de web services (API). Un SPARQL Endpoint est un service qui permet d'interroger les données structurées au format RDF. Le dépôt de ressources provenant de partenaires prioritairement issus de l'ESR ou du monde académique est également ouvert, dans le respect de [la Charte](#) définie pour Loterre. En plus de l'exposition des terminologies, le portail propose aux visiteurs des outils pour contrôler la validité d'un fichier, convertir un fichier, aligner un fichier ou annoter un texte avec une terminologie hébergée sur le portail. La première communauté des utilisateurs est celle issue de l'enseignement supérieur et de la

recherche français (ESR). Loterre peut s'adresser plus largement à toutes les communautés productrices et utilisatrices de ressources sémantiques, par exemple les traducteurs. La maintenance du portail est faite par l'Inist-CNRS. L'apport de nouveaux contenus est fait par les producteurs de terminologies.

Discussion

FooSIN : Le passage du portail TermSciences à Loterre constitue en quelque sorte une démarche de FAIRification. Quelle en a été la motivation ?

Inist-CNRS : La démarche de passage de TermSciences à Loterre a conduit de manière **indirecte à la FAIRification** des données terminologiques. Elle est au départ née **d'une volonté de publier des ressources sémantiques selon les standards du web des données ouvertes et liées (LOD)**, assortie du constat d'obsolescence technologique de la plateforme TermSciences. Dans TermSciences, les ressources étaient destinées à une consultation par l'humain. Pour Loterre, avec l'utilisation des technologies du web sémantique et des web services, ces ressources sont exploitables directement par des machines.

FooSIN : Quels étaient les impacts attendus (pour vous et pour la communauté) de ce travail de FAIRification ?

Inist-CNRS : Offrir une meilleure visibilité et permettre un plus large usage des ressources terminologiques de l'Inist et de ses partenaires, en permettant leur consultation et leur téléchargement libre sous plusieurs formats.

FooSIN : Quels ont été les moyens nécessaires pour atteindre vos objectifs ?

Inist-CNRS : Les moyens mobilisés sont issus des ressources propres de l'Inist (humaines et financières), en particulier celles du service Ingénierie terminologique mais aussi celles du Département Informatique. Des formations ont été organisées pour accompagner cette évolution du personnel : web sémantique, format RDF-SKOS, feuilles de transformation xsl, etc.

FooSIN : Quelles ont été les méthodes et les outils utilisés dans ce travail de FAIRification ?

Inist-CNRS : Nous nous sommes autant que possible appuyés sur des **outils libres** pour réaliser ce travail.

Conception de l'architecture du portail Loterre :

- pour le triplestore : [Apache Jena-Fuseki](#)
- pour permettre la consultation en ligne des ressources : [Skosmos](#)
- pour permettre à un agent logiciel distant d'interroger et de récupérer des données : [l'API REST de Skosmos](#)
- pour la formulation de requêtes SPARQL : [Sparklis](#)
- pour la partie rédactionnelle-interface web : [WordPress](#)

Transformation et enrichissement des ressources terminologiques :

- pour la gestion des ressources : Excel, [VocBench](#)
- pour l'ajout d'alignements vers d'autres référentiels : [OpenRefine](#), [OnaGUI](#)
- pour la transformation en RDF-SKOS : [Oxygen](#) (avec des feuilles de transformation conçues en interne)
- pour l'attribution de DOIs aux ressources : [Data-Cite](#)
- pour l'attribution [d'identifiants ARK](#) aux concepts : [EzARK](#)

Nous avons également souhaité partager cette expérience acquise en termes de FAIRification et d'ouverture vers le web sémantique via Loterre, au travers de [quelques services](#) proposés en ligne utiles pour contrôler la validité d'un fichier, convertir un fichier, aligner un fichier ou annoter un texte avec une terminologie hébergée sur le portail.

FooSIN : Comment les terminologies sont-elles rendues visibles et accessibles dans chacun des deux portails ?

Inist-CNRS : Dans TermSciences, les ressources sont interrogeables en ligne, sur un site web, sans restriction d'accès. Elles ne sont pas téléchargeables, le service ayant été conçu pour de la simple consultation en ligne.

Dans Loterre, les ressources sont aussi interrogeables en ligne, sur un site web, sans restriction d'accès aussi bien par un humain, via l'IHM, que par une machine en utilisant des web services ou des requêtes SPARQL. Toutes les ressources sont téléchargeables dans divers formats (SKOS-RDF, csv, pdf).

L'indexation des ressources de Loterre par google reste un sujet de préoccupation pour l'Inist. Malgré diverses actions (site map, ...), elle n'est toujours pas optimale.

FooSIN : Sous quelle(s) forme(s) les ressources sont-elles représentées et stockées initialement dans TermSciences et comment le sont-elles dans Loterre ? [I1, I2]

Inist-CNRS : Le modèle de TermSciences (base de données terminologique unique) utilise les normes ISO 16642 (TMF) et ISO 12620 (Catégories de données). Chaque entrée terminologique correspond à un concept décrit par des termes et d'autres données (définitions, contextes, etc.) issues de la fusion de plusieurs vocabulaires. Les ressources de la base de données terminologique sont stockées sous un format XML. Chaque concept possède un identifiant unique. Il n'y a pas de système de gestion des versions.

Dans Loterre, les ressources sont représentées selon le modèle de représentation *Simple Knowledge Organization System (SKOS)*. Elles sont stockées sous un format *Resource Description Framework (RDF)* dans un triplestore Apache Jena-Fuseki. Pour les métadonnées, les modèles standards utilisés sont [Dublin Core](#), [DC-terms](#), et [isothes](#). L'utilisation de modèles bien connus facilite la découverte des ressources hébergées dans Loterre. Une représentation graphique du schéma de données et de métadonnées des terminologies Loterre est proposée ici : https://www.loterre.fr/wp-content/uploads/2021/01/OntologieLoterre_bis_prod_fre.svg

FooSIN : Quelles sont les conditions de réutilisation des terminologies et sont-elles clairement définies ? Par ailleurs, les deux portails intègrent-ils un système d'authentification et la gestion des restrictions d'accès?

Inist-CNRS : Sur TermSciences, aucune mention de licence de réutilisation des terminologies n'est affichée. **Sur Loterre, la licence d'utilisation est mentionnée dans les métadonnées de chaque ressource**, aussi bien dans l'interface de consultation que dans les fichiers à télécharger. La licence doit être dans tous les cas du type « licence libre ». Sur les deux portails, la consultation est libre et peut se faire sans authentification. Les utilisateurs ne pouvant déposer eux-mêmes une ressource, cette authentification n'est en effet pas nécessaire.

FooSIN : Les fournisseurs sont-ils incités à fournir une documentation des ressources qu'ils déposent ou de les décrire dans un data paper par exemple ?

Inist-CNRS : Aucun des deux portails ne permet à un utilisateur de déposer lui-même une ressource. Dans les deux cas l'intégration est réalisée par les équipes de l'Inist, après un examen de la ressource et des échanges avec le fournisseur, qui doit fournir un certain nombre de métadonnées descriptives de la ressource et de son producteur. Pour Loterre, la liste des métadonnées nécessaires est décrite dans [la charte d'utilisation](#). Aucune autre documentation n'est requise.

Evaluation FAIR à l'aide de la grille SHARC

Afin de compléter et préciser cette interview, **FooSIN a proposé à l'Inist d'utiliser la grille SHARC produite par le groupe de travail RDA SHARC**. Il s'agit d'une grille de notation qui permet d'évaluer si on respecte les principes FAIR à travers une série de questions. Ces questions portent sur les méthodes de représentation des données, les techniques de stockage ou les moyens mis en place pour le partage. Pour en savoir plus sur la grille SHARC et la démarche d'évaluation FAIR, nous invitons les lecteurs à lire l'article « [FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles](#) ».

L'analyse des grilles est présentée et discutée plus loin dans cet article. Les deux grilles renseignées, "SHARC TermSciences" et "SHARC Loterre" sont accessibles publiquement sur le portail de données Data INRAE : <https://doi.org/10.15454/1MJDKZ>. Une représentation graphique du résultat de cette évaluation pour chacun des deux portails est proposée par l'Inist dans la Figure 1.

"FAIRitude" de TermSciences vs Loterre

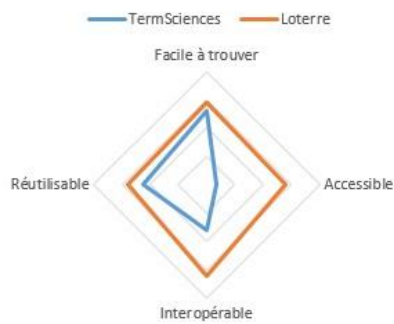


Figure 1: Synthèse des résultats de la grille SHARC sur TermSciences (bleu) et Loterre (orange)

FooSIN : De votre point de vue, quelles sont les améliorations apportées par le passage de TermSciences à Loterre vis-à-vis des principes FAIR ?

Inist-CNRS : De TermSciences à Loterre, la « FAIRitude » de la plateforme d'exposition de terminologies s'est améliorée, en particulier dans les principes **Accessible** et **Interopérable** comme illustré dans la figure. Bien que le principe **Réutilisable** ait également largement progressé, l'utilisation de la grille SHARC n'a pas permis de le montrer à sa juste mesure. Cela est sans doute dû aux difficultés rencontrées dans l'utilisation de la grille pour certains critères, car elle ne permet pas de différencier les réponses « jamais » et « non concerné ».

L'évolution constatée avec le passage de TermSciences à Loterre est tout à fait logique et n'enlève rien aux qualités de TermSciences. Il faut en effet considérer que quinze ans séparent la mise en œuvre des deux plateformes. Durant ce laps de temps, l'univers de la recherche scientifique a connu une évolution radicale en se tournant résolument vers le web sémantique, la science ouverte et la prise en compte des principes de partage et de réutilisation des données.

FooSIN : L'analyse de la grille SHARC pour Loterre permet d'identifier les points forts du portail ainsi que quelques pistes d'amélioration.

Objectif	Points forts	Améliorations possibles
Findable	<ul style="list-style-type: none"> Les terminologies sont identifiées avec un DOI (F1) Les données sont identifiées avec des URI (certaines basées sur ARK) (F1) Les métadonnées sont riches et adaptées aux ressources (F2) Les numéros de version sont indiqués (F2) 	<ul style="list-style-type: none"> Permettre une recherche sur les valeurs des métadonnées dans le portail Proposer aux producteurs un système de nommage standardisé des versions et les inciter à documenter les évolutions

Objectif	Points forts	Améliorations possibles
	<ul style="list-style-type: none"> • Les terminologies peuvent être trouvées grâce à leur catégorie thématique (F3) • Les terminologies sont indexées par les moteurs de recherche (F3) • Il existe un lien pérenne entre les métadonnées et la terminologie (URI) (F4) 	<ul style="list-style-type: none"> • Améliorer le système d'indexation / de syndication des contenus LOD dans les moteurs de recherche (contenu RDF accessible en HTML / RSS par exemple)
Accessible	<ul style="list-style-type: none"> • Les terminologies et leurs métadonnées sont enregistrées dans un triple store (A1.1) • Les terminologies sont exposées via une API REST, un SPARQL Endpoint (A1.1) • Les terminologies sont accessibles librement et pour tous (A1.2) 	<ul style="list-style-type: none"> • Rendre les anciennes versions d'une terminologie accessibles (a minima au téléchargement)
Interoperable	<ul style="list-style-type: none"> • Les métadonnées s'appuient sur des vocabulaires standards (Dublin Core principalement) et FAIR (I1, I2) • Le standard SKOS du W3C est utilisé pour toutes les terminologies (I1, I2) • Il est possible d'exprimer des liens d'équivalence (alignement) avec des éléments d'autres ressources sémantiques (I3) 	<ul style="list-style-type: none"> • Inciter les producteurs (voire des utilisateurs) de terminologies à créer des alignements avec des ressources de référence
Reusable	<ul style="list-style-type: none"> • La mention de la licence est obligatoire et représentée de manière exploitable par une application (R1.1) • Les ressources sont exportables dans plusieurs formats (PDF, RDF/XML, CSV pour les terminologies et XML, JSON-LD et Turtle pour les concepts individuels) (R1.3) • SKOS est largement connu, bien documenté (R1.3) • La documentation de Loterre recommande des outils pour 	<ul style="list-style-type: none"> • Indiquer dans quel cadre sont utilisées les terminologies • Inciter les producteurs à décrire la provenance et le processus de construction de leurs terminologies • Intégrer un outil d'aide à la création de Data Paper à partir des métadonnées et données des terminologies

Objectif	Points forts	Améliorations possibles
	manipuler les terminologies : https://www.loterre.fr/liens-utiles/ (R1.3)	

Remerciements : Le projet FooSIN remercie l'Inist, et l'équipe Loterre en particulier, de bien avoir voulu partager son expérience de FAIRification. Merci également au groupe de travail SHARC de la Research Data Alliance pour la création et la mise à disposition de la grille SHARC. Merci enfin à Romain David et Sophie Fortuno pour leur relecture attentive. Ce travail a été financé par l'ANR dans le cadre du projet FooSIN [Appel Flash science ouverte ANR-19-DATA-0019-01](#).

Citer cet article : *Majid Khayari, Véronique Reszetko, Dominique Vachez, Nathalie Vedovotto, Jérémy Yon, et al.*. De TermSciences à Loterre : comment l'Inist-CNRS a rendu les terminologies ouvertes plus conformes aux principes FAIR. 2021, [\(10.15454/djh8-yp20\)](#). [\(hal-03176063\)](#)

Références par type et dans l'ordre d'apparition dans le texte

Lectures et documents de référence

- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1). <https://doi.org/10.1038/sdata.2016.18>
- Les principes FAIR : <https://foosin.fr/2020/04/02/les-principes-fair/>
- Charte Loterre : <https://www.loterre.fr/wp-content/uploads/2020/02/Charte-Loterre.pdf>
- LOD ou web de données liées (W3C) : <https://www.w3.org/2013/data/>
- L'identifiant ARK (Archival Resource Key), consulté le 22/03/2021 <https://www.bnf.fr/fr/lidentifiant-ark-archival-resource-key>
- David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J.-E., Ioannidis, V., Bras, Y. L., Lerigoleur, E., & Cambon-Thomsen, A. (2020). FAIRness Literacy: The Achilles' Heel of Applying FAIR Principles. *Data Science Journal*, 19. <https://doi.org/10.5334/dsj-2020-032>

Portails et sites web

- Projet FooSIN : <https://foosin.fr/>
- TermSciences : <http://www.termosciences.fr/>
- Loterre : <https://www.loterre.fr/>
- Sharing Rewards and Credit WG, Research Data Alliance : <https://www.rd-alliance.org/groups/sharing-rewards-and-credit-sharc-ig>

Standards

- SKOS, Simple Knowledge Organization System : <https://www.w3.org/2004/02/skos/>
- Dublin Core expliqué par la BNF : <https://www.bnf.fr/fr/dublin-core>

Outils et services

- Apache Jena-Fuseki (Triple store) : <https://jena.apache.org/documentation/fuseki2/>
- Skosmos, outil d'exposition de thésaurus en ligne : <http://skosmos.org/> (documentation de l'API Rest de Skosmos : <http://api.finto.fi/doc/>)
- Sparklis, interface de requêtage de bases RDF : <http://www.irisa.fr/LIS/ferre/sparklis/osparklis.html>
- WordPress, logiciel de création de sites web : <https://fr.wordpress.org/download/>
- VocBench, éditeur collaboratif de thésaurus et d'ontologies : <http://vocbench.uniroma2.it/>
- OpenRefine, logiciel de transformation de données : <https://openrefine.org/>
- OnAGUI logiciel d'alignement d'ontologies : <https://github.com/lmazuel/onagui>
- Oxygen, éditeur de fichiers XML : https://www.oxygenxml.com/xml_editor/xslt_editor.html
- DataCite, service d'attribution d'identifiants DOI : <https://datacite.org/>
- eZark, service d'attribution d'identifiants ARK : <https://github.com/Inist-CNRS/ezark>
- Outils d'aide à la FAIRification Loterre : <https://www.loterre.fr/category/gerer-fr/>
- Grille SHARC, outil d'auto-évaluation pour la satisfaction des principes FAIR <https://doi.org/10.5281/zenodo.3922068>

Terminologies d'intérêt pour l'agriculture et l'alimentation publiées dans Loterre

- [Agriculture et systèmes d'élevage](#)
- [Biotechnologies de la reproduction](#)
- [Écologie microbienne du tube digestif](#)
- [Glossaire de génétique moléculaire et de génie génétique](#)
- [La cuisson-extrusion](#)
- [Lexique de pédologie](#)
- [L'ionisation dans l'industrie agro-alimentaire](#)
- [Techniques de séparation par membranes](#)
- [Technologie des semences](#)
- [Vocabulaire de la Politique Agricole Commune](#)