



HAL
open science

3 outils pour transformer du tabulé en SKOS

Sophie Aubin, Jérémy Yon

► **To cite this version:**

| Sophie Aubin, Jérémy Yon. 3 outils pour transformer du tabulé en SKOS. 2022. hal-03852086

HAL Id: hal-03852086

<https://hal.science/hal-03852086>

Submitted on 14 Nov 2022

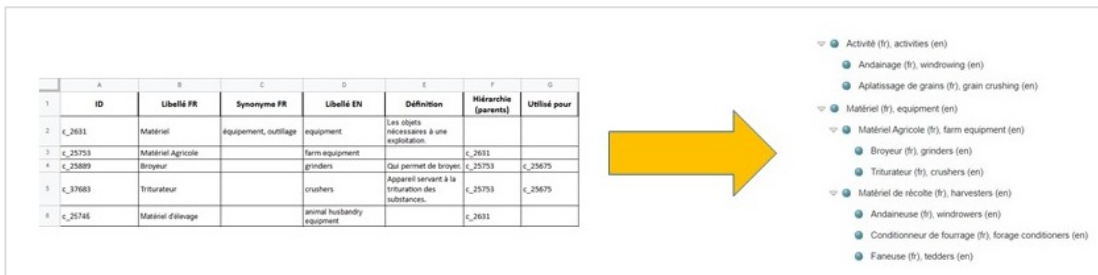
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

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.

3 outils pour transformer du tabulé en SKOS



Cette ressource s'adresse aux personnes qui souhaitent **adopter le standard SKOS (Simple Knowledge Organization System) pour représenter une ressource terminologique** (glossaire, thésaurus...) initialement représenté dans un format tabulé (csv, tsv, xls...). Les motivations peuvent être diverses : utiliser des outils métiers s'appuyant sur SKOS pour visualiser ou éditer le vocabulaire, l'exposer sur un portail publique comme Agroportal ou Loterre ou, de manière plus générale, pour rendre la ressource utilisable dans un système d'information ou par d'autres utilisateurs, en d'autres termes pour le rendre plus FAIR : plus Facile à trouver (Findable), Accessible, Interopérable, Réutilisable (Reusable).



SKOS est un modèle relativement simple et facile à s'approprier pour les personnes familières des technologies web sémantique et en particulier de RDF. Pour les personnes moins aguerries, **le choix de l'outil de transformation aura une incidence sur le temps de réalisation et la qualité du résultat**. Le projet FooSIN et le service Vocabulaires Ouverts ont testé et comparé 3 outils avec comme objectif de guider ce choix en fonction du profil de l'utilisateur et de la complexité des données à transformer.

Nous avons choisi de tester **SKOS Play! Convertir**, **OpenRefine** et **Sheet2RDF intégré à VocBench** car ils ne nécessitent pas de compétences informatiques avancées pour les

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

1. une introduction au modèle SKOS et une analyse de ses avantages pour l'application des principes FAIR (page séparée) ;
2. une courte présentation de chacun des 3 outils : SKOS Play!, OpenRefine et Sheet2RDF
3. une procédure pas à pas à partir d'un exemple simple pour transformer du format tabulé en SKOS avec chacun des trois outils :
 - Pas à pas avec **SKOS Play**.
 - Pas à pas avec **OpenRefine**.
 - Pas à pas avec **Sheet2RDF dans VocBench**.
4. une présentation des avantages et inconvénients de chacune des trois solutions testées.

1. SKOS : un standard pour une ressource sémantique simple et FAIR

Accéder à la section

2. Des outils pour transformer un fichier tabulé au format SKOS

Nous avons choisi de tester **trois outils libres et gratuits** qui permettent de transformer un fichier tabulaire contenant un vocabulaire simple en une ressource SKOS. Les trois outils sont **SKOS PLAY**, **OpenRefine** et **VocBench-Sheet2RDF**.

SKOS Play! Convertir

SKOS Play est un ensemble de services en ligne développés et mis à disposition gratuitement par la société Sparna pour travailler avec des ressources sémantiques au format SKOS. En plus de proposer de nombreux modes de visualisation de données SKOS, SKOS Play! propose des services très pratiques comme « vérifier le respect des règles d'intégrité du modèle SKOS » et aussi « convertir une feuille de calcul Excel en fichier SKOS ». C'est le service "Convertir", disponible à l'adresse url suivante : <http://labs.sparna.fr/skos-play/convert> qui nous intéresse.

SKOS Play !

Accueil ▶ Play! Convertir Tester A propos Forum fr ▼

Où se trouvent les données excel à convertir ?

Dans un des fichiers d'exemple fourni [Télécharger l'exemple : Exemple 1 \(simple exemple, in english\)](#)

Dans un fichier sur mon ordinateur

(Extensions supportées: .xls ou .xlsx. Les fichiers OpenOffice ne fonctionnent pas !)

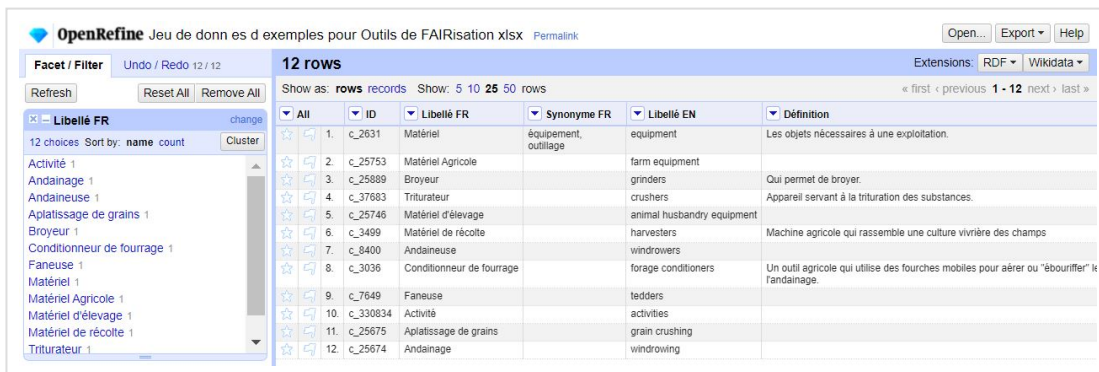
Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER

ordonner et aussi les transformer d'un format à un autre. Il est facile à installer sur un poste de travail sous Windows, Mac ou Linux. OpenRefine n'est pas un outil centré autour du modèle de données SKOS comme peuvent l'être les deux autres. Pour l'usage que nous en faisons, il est nécessaire d'installer l'[extension RDF-openRefine](#). La [documentation](#), en anglais, est riche et pratique. [Télécharger OpenRefine](#) et accéder à de nombreuses extensions.



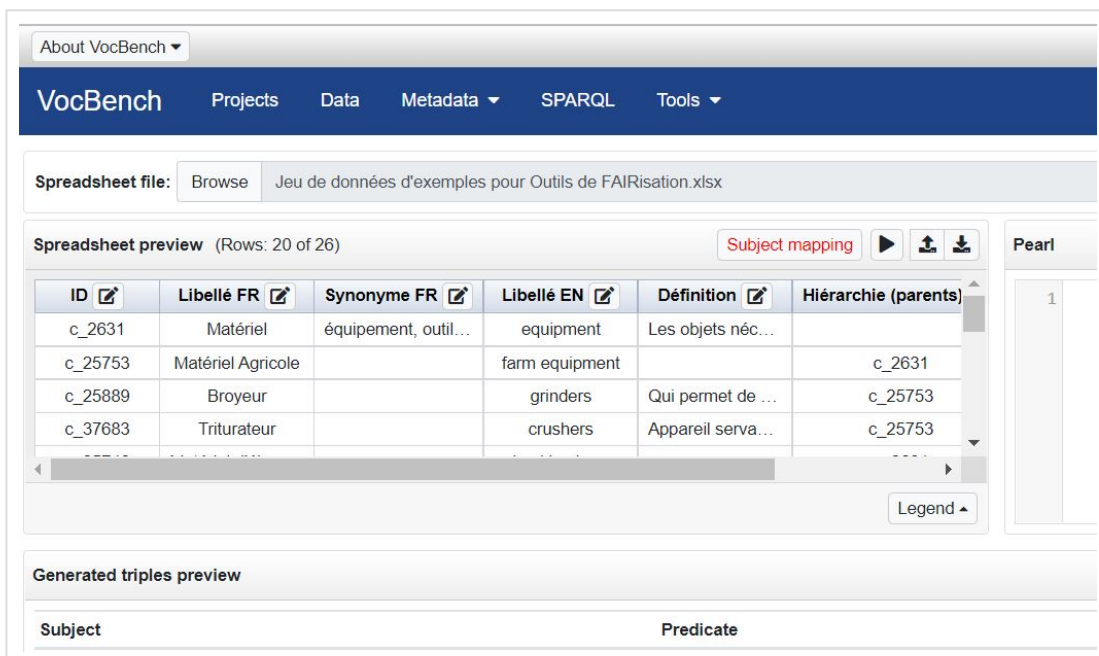
The screenshot shows the OpenRefine web interface. At the top, it says "OpenRefine Jeu de données d'exemples pour Outils de FAIRisation.xlsx". Below that, there's a "Facet / Filter" section with "Libellé FR" selected. The main table has 12 rows. The columns are: All, ID, Libellé FR, Synonyme FR, Libellé EN, and Définition. The data includes various agricultural equipment like "Matériel", "Matériel Agricole", "Broyeur", "Triturateur", etc.

All	ID	Libellé FR	Synonyme FR	Libellé EN	Définition
1.	c_2631	Matériel	équipement, outillage	equipment	Les objets nécessaires à une exploitation.
2.	c_25753	Matériel Agricole		farm equipment	
3.	c_25889	Broyeur		grinders	Qui permet de broyer.
4.	c_37683	Triturateur		crushers	Appareil servant à la trituration des substances.
5.	c_25746	Matériel d'élevage		animal husbandry equipment	
6.	c_3499	Matériel de récolte		harvesters	Machine agricole qui rassemble une culture vivrière des champs
7.	c_8400	Andaineuse		windrowers	
8.	c_3036	Conditionneur de fourrage		forage conditioners	Un outil agricole qui utilise des fourches mobiles pour aérer ou "bouriffer" le fandrainge.
9.	c_7649	Faneuse		tedders	
10.	c_330834	Activité		activities	
11.	c_25675	Aplattissage de grains		grain crushing	
12.	c_25674	Andainage		windrowing	

Sheet2RDF dans VocBench

VocBench est un outil d'édition de thésaurus ou d'ontologies collaboratif et open source développé par l'Université Tor Vergata, Rome. Le logiciel peut être installé sur un serveur ou sur un poste de travail. En plus des fonctionnalités d'édition, VocBench embarque divers outils pour vérifier la syntaxe d'un fichier, valider des alignements entre vocabulaires ou transformer un fichier tabulaire en RDF. C'est ce dernier outil appelé **Sheet2RDF** que nous testons.

[Télécharger VocBench](#) et accéder à la [documentation](#).



The screenshot shows the VocBench web interface. At the top, there's a navigation bar with "VocBench", "Projects", "Data", "Metadata", "SPARQL", and "Tools". Below that, there's a "Spreadsheet file" section with a "Browse" button and the file "Jeu de données d'exemples pour Outils de FAIRisation.xlsx". The "Spreadsheet preview" section shows a table with 20 rows (out of 26) and columns: ID, Libellé FR, Synonyme FR, Libellé EN, Définition, and Hiérarchie (parents). The data is the same as in the OpenRefine screenshot. There's also a "Subject mapping" button and a "Legend" button. At the bottom, there's a "Generated triples preview" section with columns for "Subject" and "Predicate".

ID	Libellé FR	Synonyme FR	Libellé EN	Définition	Hiérarchie (parents)
c_2631	Matériel	équipement, outillage	equipment	Les objets néc...	
c_25753	Matériel Agricole		farm equipment		c_2631
c_25889	Broyeur		grinders	Qui permet de ...	c_25753
c_37683	Triturateur		crushers	Appareil serva...	c_25753

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER

Présentation du jeu de données pour l'exercice

Ce jeu de données test est un ensemble de 11 concepts décrivant du matériel agricole (inspirés d'Agrovoc) et représentés sous forme d'un tableau dont un extrait est montré dans le ci-dessous. Le jeu de données test est téléchargeable au **format Excel** (utilisé pour les tests) ou **texte tabulé** (format ouvert).

	A	B	C	D	E	F	G
1	ID	Libellé FR	Synonyme FR	Libellé EN	Définition	Hiérarchie (parents)	Utilisé pour
2	c_2631	Matériel	équipement, outillage	equipment	Les objets nécessaires à une exploitation.		
3	c_25753	Matériel Agricole		farm equipment		c_2631	
4	c_25889	Broyeur		grinders	Qui permet de broyer.	c_25753	c_25675
5	c_37683	Triturateur		crushers	Appareil servant à la trituration des substances.	c_25753	c_25675
6	c_25746	Matériel d'élevage		animal husbandry equipment		c_2631	

Dans le tableau suivant, nous indiquons pour chaque colonne du tableau d'origine une description ainsi qu'une explication de la manière dont l'information contenue sera transposée en SKOS.

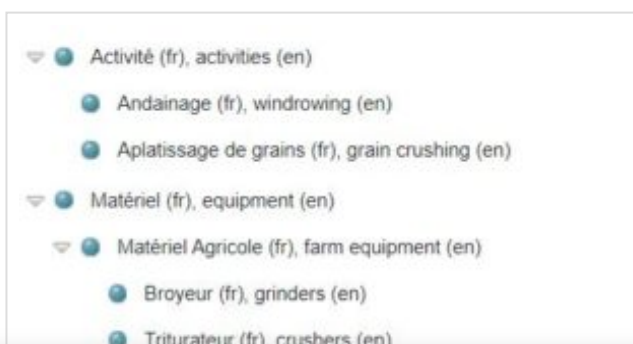
Nom de colonne	Description	Exemple	Correspondance en RDF/SKOS
ID	un identifiant de concept	<i>c_2631</i>	Cette valeur sera transformée en URI en le concaténant avec l'URI de base du vocabulaire : <i>https://opendata.inrae.fr/Thesaurus/</i> . Ainsi l'URI du concept décrit dans la ligne 2 sera <i>https://opendata.inrae.fr/Thesaurus/c_2631</i> et pourra être raccourcie en <i>th:c_2631</i> puisque nous avons déclaré le préfixe <i>th</i> pour cet URI de base.
Libellé FR	un terme préférentiel français	<i>Matériel</i>	Cette valeur deviendra l'objet de la propriété skos:prefLabel . Pour rappel, cette propriété associe au concept un terme préféré dans chaque langue. Dans notre cas, il s'agira d'un littéral pour lequel on précisera qu'il s'agit du français à l'aide de l'attribut <i>fr</i>
Synonyme	une liste de	<i>équipement,</i>	Ces valeurs deviendront chacune l'objet

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

	anglais		l'objet de la propriété skos:prefLabel . La langue du synonyme sera spécifiée par l'attribut <i>en</i> pour anglais
Définition	une définition en français	<i>Les objets nécessaires à une exploitation</i>	Cette valeur sera transformée pour être l'objet de la propriété skos:definition . Pour rappel, cette propriété associe au concept une définition textuelle. On peut préciser la langue, ici on spécifiera <i>fr</i> pour français
Hiérarchie (parents)	l'identifiant du concept plus générique s'il existe	<i>c_25753</i>	Cette valeur deviendra l'objet de la propriété skos:broader . Pour rappel, cette propriété associe au concept un concept parent. La valeur attendue est un URI de concept défini dans une autre ligne
Utilisé pour	l'identifiant d'un concept lié par une relation spécifique : un matériel agricole est utilisé pour effectuer une tâche représentée par un autre concept	<i>c_25675</i>	La valeur de cette colonne sera transformée pour être l'objet de la propriété <i>th:Utiliser_pour</i> . Le modèle de données SKOS ne possède pas de propriété pour préciser ce lien. Nous définissons donc une nouvelle propriété adhoc intitulée <i>th:Utiliser_pour</i> .

Nous supposons que tous les concepts présents dans le fichier appartiennent au même schème. Ce schème a comme URI <http://opendata.inrae.fr/Thesaurus/ThesaurusExemple>. En SKOS, on utilisera la propriété *skos:inScheme*. La propriété *skos:hasTopConcept* sera également utile pour indiquer quels concepts se trouvent "en haut" de l'arbre, i.e. les concepts qui n'ont pas de parents.

Après transformation, l'arbre des concepts se présente ainsi :



Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER

Et voici un extrait du fichier SKOS en RDF/XML.

```
<rdf:Description rdf:about="http://opendata.inrae.fr/Thesaurus/c_2631">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:prefLabel xml:lang="fr">Matériel</skos:prefLabel>
  <skos:prefLabel xml:lang="en">equipment</skos:prefLabel>
  <skos:definition xml:lang="fr">Les objets nécessaires à une exploitation.</skos:definition>
  <skos:inScheme>http://opendata.inrae.fr/Thesaurus/ThesaurusExemple</skos:inScheme>
</rdf:Description>

<rdf:Description rdf:about="http://opendata.inrae.fr/Thesaurus/c_25753">
  <rdf:type rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>
  <skos:prefLabel xml:lang="fr">Matériel Agricole</skos:prefLabel>
  <skos:prefLabel xml:lang="en">farm equipment</skos:prefLabel>
  <skos:inScheme>http://opendata.inrae.fr/Thesaurus/ThesaurusExemple</skos:inScheme>
</rdf:Description>

<rdf:Description rdf:about="http://opendata.inrae.fr/Thesaurus/c_2631">
  <skos:narrower rdf:resource="http://opendata.inrae.fr/Thesaurus/c_25753"/>
</rdf:Description>

<rdf:Description rdf:about="http://opendata.inrae.fr/Thesaurus/c_25753">
  <skos:broader rdf:resource="http://opendata.inrae.fr/Thesaurus/c_2631"/>
</rdf:Description>
```

Détail des tests en pas à pas

A partir du jeu de données test, nous vous indiquons la procédure pour transformer un fichier tabulé en SKOS avec chacun des 3 outils testés :

- Lire le pas à pas avec [SKOS Play](#).
- Lire le pas à pas avec [OpenRefine](#).
- Lire le pas à pas avec [Sheet2RDF dans VocBench](#).

4. Comparatif des 3 outils

Nous synthétisons les résultats de nos tests pour aider le lecteur à choisir la solution la mieux adaptée à sa situation. En cliquant sur l'intitulé de la ligne, vous pouvez accéder à une analyse plus détaillée.

	SKOS PLAY	Sheet2RDF/VocBench	OpenRefine
	Convertir		
format d'entrée	classeur Excel	classeur Excel, LibreOffice ou OpenOffice fichier CSV et TSV	classeur Excel CSV et TSV
format de sortie	Turtle, RDF/XML, N3, TriG, N-Triples, N-Quads	RDF/XML, RDF/JSON, N3, TriG, N-Triples, N-Quads, JSON-LD	Turtle, RDF/XML

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER

cellules multivaluées		Doivent être scindées au préalable	
création automatique de schème	Oui	Non	Non
	Un schème par feuillet dans le classeur	Création manuelle ou doivent être inclus dans le jeu de données	Création manuelle dans le <i>skeleton rdf</i> ou doivent être inclus dans le jeu de données
calcul des top concepts	Oui	Non	Non
		Doivent être indiqués dans le jeu de données	Doivent être indiqués dans le jeu de données
génération automatique des URI pour SKOS-XL	Oui	Oui	Pas facilement L'utilisateur doit construire sa propre stratégie
documentation	Documentation en anglais, claire et bien illustrée	Documentation en anglais mais très légère sur la partie utilisation de l'application Sheet2RDF et difficile à comprendre pour un non-informaticien	Documentation officielle en anglais. Pas de documentation en ligne pour l'extension rdf

Quel que soit l'outil que vous aurez choisi, nous vous recommandons de tester le fichier produit à l'aide d'un outil de validation de SKOS comme [SKOS Play Tester](#) ou [VocBench Integrity Constraint Validator](#). Les tests permettent de détecter des erreurs ou des informations manquantes.

Cette ressource a été développée dans le cadre du projet ANR FooSIN ([ANR-19-DATA-0019-](#)

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER



Auteurs (INRAE): Jérémy Yon, Sophie Aubin

Relecteurs (INRAE) : Catherine Roussey, Sylvie Cocaud, Magalie Weber

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER

Résultats du comparatif Skos Play!, OpenRefine et Sheet2RDF

Installation des outils

SKOS Play Convertir est disponible en ligne ce qui facilite son utilisation car il n'y a pas d'installation à faire. Cependant, vous êtes obligés d'envoyer vos données chez un tiers.

OpenRefine et **Vocbench Sheet2RDF** sont des applications qui fonctionnent comme un petit serveur Web sur votre propre ordinateur. Il suffit de télécharger un dossier contenant l'application, de le décompresser et de double-cliquer sur l'icône de l'application pour la démarrer. Leur utilisation demande d'avoir JAVA JRE/JDK sur sa machine (taper `java -version` dans la console de commande pour savoir s'il est installé). Pour OpenRefine, il est nécessaire de télécharger et de placer l'extension **RDF-OpenRefine** dans le répertoire adéquat en suivant la procédure décrite sur la [page de téléchargement](#).

Formats d'entrée

SKOS Play Convertir ne prend en charge que les formats *xls* ou *xlsx*.

Vocbench Sheet2RDF accepte les classeurs des trois licences (Excel, OpenOffice et LibreOffice) et des fichiers CSV et TSV.

OpenRefine accepte les classeurs des trois licences (Excel, OpenOffice et LibreOffice) et des fichiers tabulés avec les séparateurs classiques (tabulation, virgule) ou autre (paramétrable).

[Retour au tableau](#)

Formats de sortie

SKOS Play Convertir permet d'exporter le résultat en RDF avec plusieurs formats différents (Turtle, RDF/XML, N3, TriG, N-Triples, N-Quads)

VocBench Sheet2RDF permet d'exporter le résultat en RDF dans plusieurs formats dont JSON.

OpenRefine permet d'exporter le résultat en RDF suivant les deux formats RDF/XML et Turtle.

Pour en savoir plus sur les différentes sérialisations de RDF :

- Inria. (2015, 2 mars). *Syntaxes de sérialisation*. [Vidéo]. Canal-U. <https://www.canal-utv/84513>. (Consultée le 25 mars 2022)

Capacité à représenter la richesse données

Les trois outils nous laissent la possibilité d'inclure des propriétés n'appartenant pas au modèle SKOS.

Seuls **SKOS Play** et **OpenRefine** prennent en charge les cellules multivaluées.

[Retour au tableau](#)

Assistance pour travailler avec le modèle de données SKOS et le langage RDF

SKOS Play Convertir propose des options afin de réaliser des opérations à notre place. L'outil déclare automatiquement un schème. Si vous avez plusieurs schèmes au sein de votre jeu de données, il vous suffit de les séparer dans différentes feuilles de calcul. L'outil lira toutes les feuilles et créera un seul fichier sortie. Il calcule automatiquement les *top concepts* de votre vocabulaire. De plus, SKOS Play Convertir propose de créer automatiquement les libellés SKOS-XL si le fichier source contient des informations sur les termes (en savoir plus sur [SKOS-XL](#)). La seule fonction non présente que nous aimerions avoir est une option pour créer une propriété inverse à une propriété présente dans le jeu de données.

VocBench Sheet2RDF propose une méthode de configuration plus experte que les deux autres outils. On peut utiliser des propriétés SKOS ou des propriétés issues d'autre modèle de données. VocBench Sheet2RDF permet d'utiliser le modèle de données SKOS-XL et fournit des mécanismes pour générer des URIs. L'outil présente toutefois quelques limites. Aucune aide n'est proposée pour la création du schème, il faut donc prévoir cette propriété dans le jeu de données source. L'utilisateur devra également s'assurer que son jeu de données ne contient pas de cellules multivaluées, Sheet2RDF ne proposant rien pour les traiter.

OpenRefine n'intègre pas d'aide spécifique pour SKOS ou SKOS-XL. Il est donc nécessaire de bien connaître le modèle de données SKOS pour utiliser OpenRefine.

[Retour au tableau](#)

Documentation

SKOS Play Convertir : la documentation de l'outil est visible sous la zone de chargement des données et est très bien illustrée.

VocBench Sheet2RDF : il n'y a pas d'aide contextuelle dans Sheet2RDF mais plusieurs pages de [documentation](#) qui reste cependant assez technique et n'existe qu'en anglais. Un tutoriel est [disponible](#) mais il ne couvre pas complètement l'utilisation de l'outil.

OpenRefine : la procédure pour créer les correspondances entre les colonnes du tableau et les propriétés SKOS est bien décrite dans la [documentation](#) qui reste cependant générique. Il n'ex

Cette ressource a été développée dans le cadre du projet ANR FooSIN (ANR-19-DATA-0019-01)



Auteurs (INRAE): Jérémy Yon, Sophie Aubin

Relecteurs (INRAE) : Catherine Roussey, Sylvie Cocaud, Magalie Weber

Transformer un fichier tabulé en SKOS avec OpenRefine



Cette documentation « pas à pas » vous montre comment utiliser **OpenRefine** pour transformer un vocabulaire au format tabulé en une ressource représentée à l'aide du modèle standard SKOS. Il permet d'aborder quelques fonctionnalités de base d'OpenRefine, d'expérimenter l'extension RDF et de se familiariser avec le modèle SKOS.

Cette page est un des résultats du [test comparatif de 3 outils pour transformer un vocabulaire au format tabulé vers SKOS](#) (projet FooSIN). Toutes [les données sont publiées](#) sur l'entrepôt ouvert [recherche.data.gouv](https://recherche.data.gouv.fr).

Libellé FR	ID	Libellé FR	Synonyme FR	Libellé EN	Définition
	1	c_2631	Matériel	équipement, outillage	Les objets nécessaires à une exploitation.
Activité 1	2	c_26753	Matériel Agricole	farm equipment	
Andaineuse 1	3	c_25889	Broyeurs	grinders	Qui permet de broyer.
Andaineuse 1	4	c_37683	Triturateur	crushers	Appareil servant à la trituration des substances.
Aplatisseage de grans 1	5	c_25746	Matériel d'élevage	animal husbandry equipment	

OpenRefine est un outil open source pour travailler à partir de jeux de données mal ou peu structurées pour les nettoyer, les ordonner et aussi les transformer d'un format à un autre (beaucoup de possibilités). Vous pouvez le télécharger et l'installer sur votre poste. Il s'agit d'un outil générique, pas spécifique au modèle SKOS ni même au langage RDF. Pour ce test, nous avons utilisé la Version 3.5.2 d'OpenRefine. Il vous faudra d'ailleurs installer l'extension [RDF extension](#) suivant les instructions disponibles [ici](#).

Présentation du jeu de test

Le fichier source est un fichier tabulé au format .xlsx (lien vers le fichier [Jeu de données – entree OpenRefine.xlsx](#)). D'autres formats tabulés .xls, .csv ou .ods sont possibles. Aucune structure particulière n'est attendue par OpenRefine, c'est vous qui allez déclarer les correspondances entre vos colonnes et les propriétés du modèle SKOS. Nous vous conseillons de garder une structure simple avec une colonne par propriété (URI, label, etc.) et une ligne par concept. Il est possible de traiter des cellules multivaluées à condition qu'il y ait un séparateur fixe entre les différentes valeurs. Bien nommer vos colonnes rendra la définition des correspondances vers SKOS plus facile. La ligne 1 dans la Figure ci-dessous est donc une ligne facultative mais recommandée.

	A	B	C	D	E	F	G
1	ID	Libellé FR	Synonyme FR	Libellé EN	Définition	Hierarchie (parents)	Utilisé pour
2	c_2631	Matériel	équipement, outillage	equipment	Les objets nécessaires à une exploitation.		
3	c_25753	Matériel Agricole		farm equipment		c_2631	
4	c_25889	Broyeur		grinders	Qui permet de broyer.	c_25753	c_25675
5	c_37683	Tritrateur		crushers	Appareil servant à la trituration des substances.	c_25753	c_25675
6	c_25746	Matériel d'élevage		animal husbandry equipment		c_2631	

Chargez le fichier

Le chargement du fichier se fait depuis l'onglet « Create Project » (encadré jaune dans la figure ci-dessous). Dans la plupart des cas, OpenRefine détecte automatiquement le format du fichier en entrée (encadré rouge). Il vous présente ensuite une prévisualisation du jeu de données (encadré vert). Vous pouvez vérifier que le chargement a été fait correctement, ajuster les options si besoin, notamment pour l'encodage du fichier. Cliquez sur « Create Project » en haut à droite pour créer votre projet.

Dans un premier temps, l'interface vous permet de vérifier, corriger et filtrer vos données si vous en avez besoin. OpenRefine offre de nombreuses fonctionnalités, voir la documentation en ligne. Une fois la visualisation de votre jeu de données conforme à vos attentes, vous allez définir la correspondance entre vos colonnes et les éléments du modèle SKOS.

Accédez au module de correspondances

Cliquez sur le bouton « RDF » (cadre rouge ci-dessus) et choisissez « Edit RDF skeleton » qui ouvre la page de « *RDF Schema alignment*. A l'aide de cette interface, vous allez

déclarer les règles de construction des triplets RDF qui constitueront votre ressource en SKOS. Pendant tout le travail, pensez à sauvegarder régulièrement votre “skeleton” (bouton Save en bas à droite). Pensez à utiliser la fonctionnalité “Preview” disponible à différents endroits pour contrôler le résultat de vos actions.

RDF Schema alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: http://127.0.0.1:3333/ Edit

RDF skeleton | RDF Preview

Available prefixes: rdf owl rdfs +Add Manage

(Row index) URI X> property?> ID Cell
Add type X> property?> Libellé FR Cell
 X> property?> Synonyme FR Cell
 X> property?> Libellé EN Cell
 X> property?> Définition Cell
 X> property?> Hiérarchie (parents) Cell
 X> property?> Utilisé pour Cell
 Add property

Add another root node

Save

OK Cancel

Déclarez l'URI de base et les préfixes

La première chose à faire est de renseigner l'URI de base de votre ressource (encadré rouge de la figure ci-dessus), dans notre exemple *http://opendata.inrae.fr/Thesaurus/*. L'URI de base peut se terminer par / ou bien #. Elle servira à construire les URI pour votre vocabulaire.

Ensuite, vous devez déclarer les schémas RDF que vous souhaitez utiliser (encadré vert), ainsi qu'un préfixe pour chacun. Les schémas les plus courants sont présents par défaut. Pour ajouter un nouveau préfixe, cliquez sur +Add (encadré vert ci-dessus). Pour notre cas, nous avons besoin en plus de déclarer SKOS avec les valeurs suivantes:

Add new prefix

Prefix: skos

URI: http://www.w3.org/2004/02/skos/core#
(a suggestion by prefix.cc is provided)

OK Cancel Vocabulary import...

Si vous utilisez des éléments d'autres schémas, pensez à les déclarer également. Par exemple, dans notre jeu de données, on crée le préfixe *th* associé à l'URI *http://opendata.inrae.fr/Thesaurus/* pour pouvoir créer une propriété adhoc (idéalement, cette propriété doit être définie dans une ontologie). Si le schéma n'est pas accessible en ligne, sélectionnez l'option “Add prefix only” dans le menu “Vocabulary import”.

Ci-dessous, le résultat du paramétrage pour notre exemple :

Base URI: <http://opendata.inrae.fr/Thesaurus/> Edit

RDF skeleton

RDF Preview

Available prefixes: rdf owl skos rdfs th +Add Manage

Déclarez les patrons de correspondance

Vient ensuite le moment de créer les correspondances entre la structure de votre fichier en colonnes et les éléments du modèle SKOS. Pour chaque patron de triplet à générer, vous allez indiquer où trouver le sujet (1), la propriété SKOS à utiliser (2) et où trouver l'objet du prédicat (3). Notez que les intitulés de vos colonnes n'apparaissent plus dans la partie "objet" du prédicat comme précédemment. Il s'agit de toute évidence d'un bug. Vous devrez donc aller les sélectionner un à un.

RDF Schema alignment

The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.

Base URI: <http://opendata.inrae.fr/Thesaurus/> Edit

RDF skeleton

RDF Preview

Available prefixes: th rdf owl skos rdfs +Add Manage

(Row index) URI
Add type

Add property

1

2

3

Déclarez le sujet des triplets (concept)

Commençons par définir le sujet du triplet. Il doit être un identifiant de concept. Cliquez sur (Row index) URI (repère 1 dans la figure ci-dessus) pour ouvrir la fenêtre de paramétrage du sujet (Fenêtre *RDF node* ci-contre). Par

RDF node

Use content...

- (Row index)
- ID
- Libellé FR
- Synonyme FR
- Libellé EN
- Définition
- Hiérarchie (parents)
- Utilisé pour
- Constant value

Content used ...

- URI
- Text
- Language
- Integer
- Non-integer
- Date (YYYY-MM-DD)
- Date/time (YYYY-MM-DD HH.MM.SS)
- Boolean
- Custom (specify type URI)
- Blank

Use expression...

value

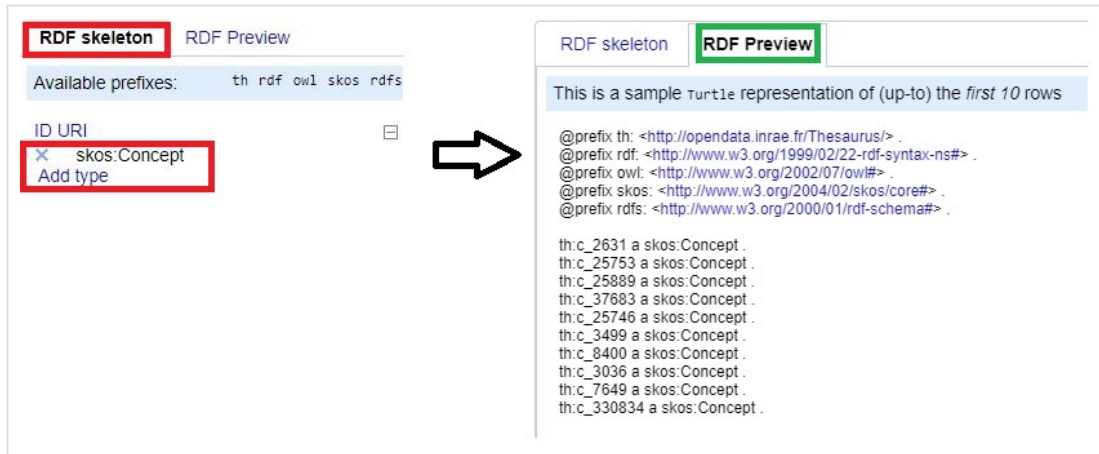
Preview edit

OK Cancel

défaut, l'outil propose que ce soit le numéro de ligne mais comme notre fichier contient déjà des identifiants (colonne A), c'est cette valeur qu'il faut récupérer. Dans la boîte *Use content...*, cochez alors *ID* (intitulé de la colonne A) au lieu de *(Row index)*.

Dans la boîte *Content used...*, indiquez qu'il s'agit d'un URI. L'URI sera construit en utilisant la "Base URI" que vous avez définie auparavant : [Base URI + valeur de la colonne ID]. Cliquez sur ok pour retourner à la fenêtre *RDF Schema alignment*.

Ensuite, utilisez la fonction qui permet de donner un « rdf:type » à l'URI en cours de création. Pour cela, il faut cliquer sur *Add type* et ajouter le type que l'on veut, dans notre cas, ce sera *skos:Concept*. Vous pouvez voir un premier résultat de transformation de vos données dans l'onglet RDF Preview.

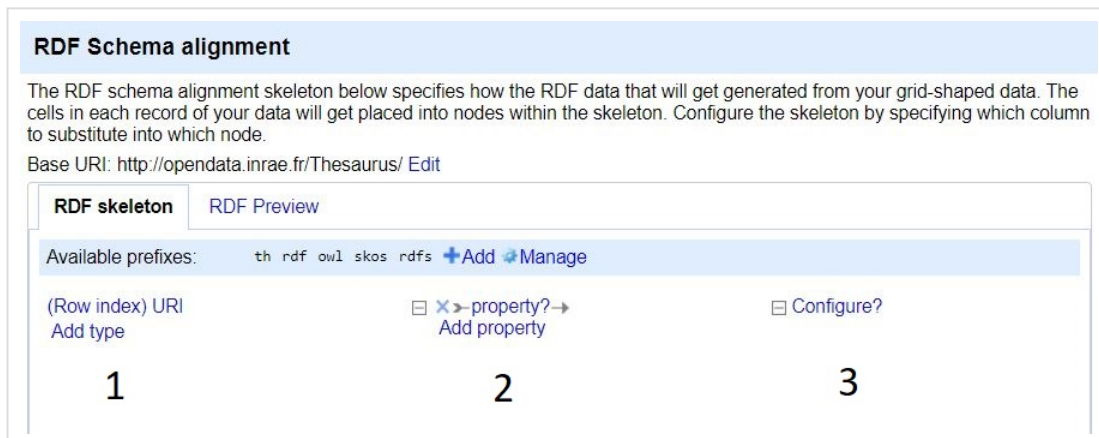


The screenshot shows two panels. The left panel, titled 'RDF skeleton', has a tab 'RDF Preview' selected. It shows 'Available prefixes: th rdf owl skos rdfs'. Below, there is a table with columns 'ID' and 'URI'. The row for 'skos:Concept' is selected, and the 'Add type' button is highlighted. An arrow points to the right panel, titled 'RDF Preview', which has a tab 'RDF Preview' selected. It shows a sample Turtle representation of the data, including prefixes and a list of concepts like 'th:c_2631 a skos:Concept'.

Déclarez les prédicats et les objets

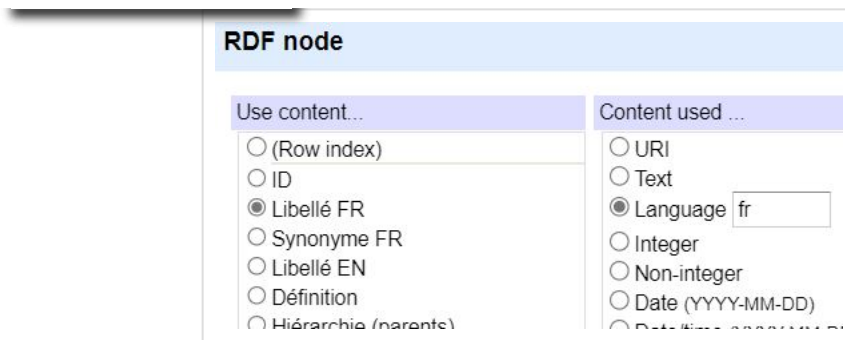
Maintenant que le sujet est paramétré, il faut déclarer l'objet du triplet ainsi que le prédicat.

L'**objet** peut être une valeur littérale avec ou sans langue, un entier, une date... pour les propriétés de type "*data property*" ou l'URI d'une autre ressources RDF pour les propriétés de type "*object property*" (voir l'[introduction à RDF](#)).

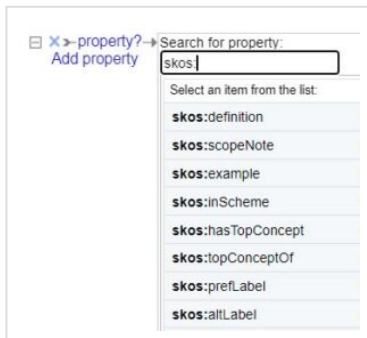


The screenshot shows the 'RDF Schema alignment' window. It has a title bar 'RDF Schema alignment' and a description: 'The RDF schema alignment skeleton below specifies how the RDF data that will get generated from your grid-shaped data. The cells in each record of your data will get placed into nodes within the skeleton. Configure the skeleton by specifying which column to substitute into which node.' Below the description is the 'Base URI: http://opendata.inrae.fr/Thesaurus/ Edit'. There are two tabs: 'RDF skeleton' and 'RDF Preview'. The 'RDF skeleton' tab is active. It shows 'Available prefixes: th rdf owl skos rdfs +Add +Manage'. Below, there is a table with columns '(Row index) URI', 'Add type', 'Add property?', and 'Configure?'. The row for '1' is selected, and the 'Add property?' button is highlighted. The 'Configure?' button is also visible.

Toujours dans la fenêtre *RDF schema alignment*, cliquez sur [Add property?](#) (repère 2). Un nouveau champ, *Configure?*, apparaît à droite (repère 3). Cliquez dessus pour paramétrer l'objet de votre triplet. La fenêtre de paramétrage (RDF Node) est la même que pour le sujet. Dans la boîte *Use content*, sélectionnez une colonne de votre fichier. Déclarons par exemple *Libellé FR*. Ici, pour indiquer qu'il s'agit d'une valeur littérale en français, il faut sélectionner Language et indiquer "fr" dans le champ libre.

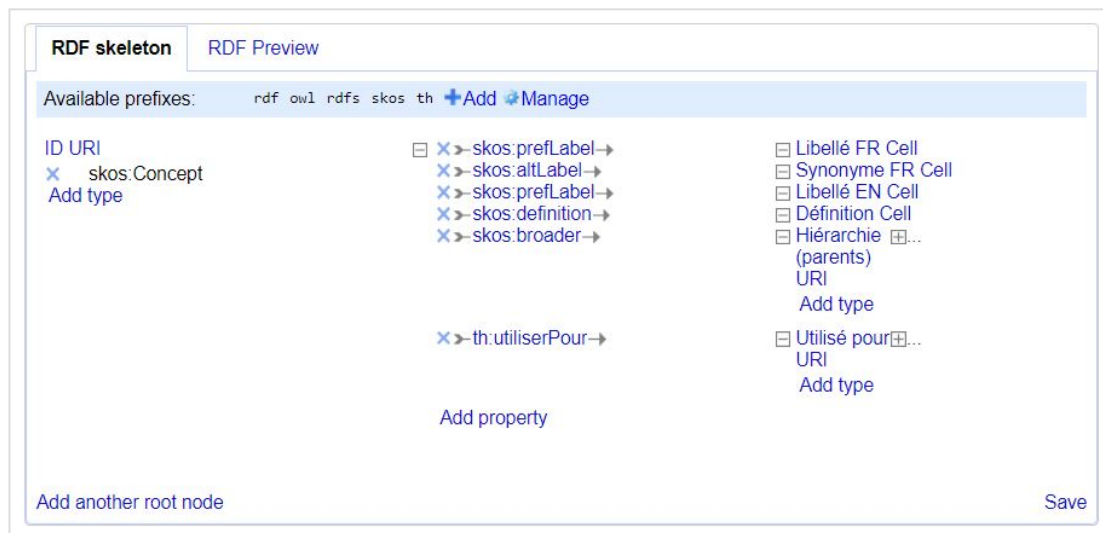


Si vos cellules sont multivaluées, voir le paragraphe suivant. Cliquez sur Ok pour revenir à la fenêtre *RDF Schema alignment*.



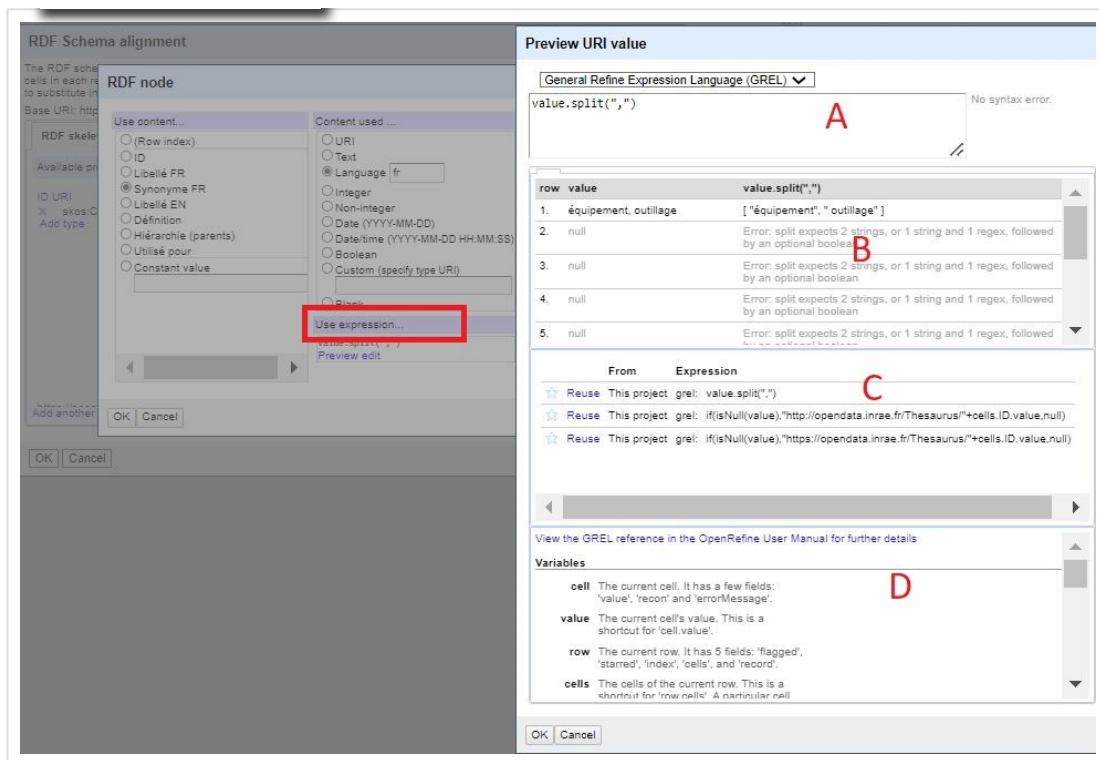
Pour paramétrer le **prédicat**, cliquez sur *Add property?*. Saisissez le préfixe *skos* et choisissez la propriété SKOS adéquate (*skos:prefLabel* dans notre exemple). Pour la propriété adhoc *th:utiliserPour*, indiquez simplement le préfixe *th* et le nom de propriété.

Répétez ces opérations pour chaque propriété à construire à partir des colonnes de votre fichier. Vous devez obtenir un ensemble de correspondances comme celui montré ci-dessous.



Cas particulier des cellules multivaluées

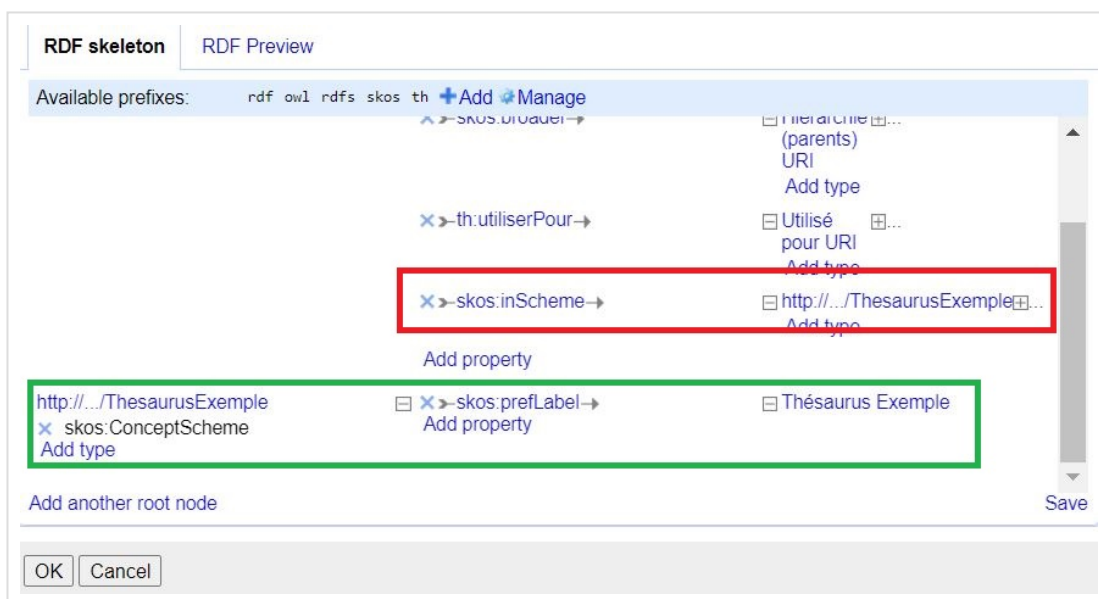
Si votre fichier présente des cellules multivaluées (les *Synonymes FR* séparés par une virgule dans notre exemple), utilisez la boîte *Use expression* comme dans la figure ci-dessous. Le module contient une fenêtre d'édition (A), une prévisualisation du résultat (B), un historique (C) et de la documentation des fonctions disponibles (D). En insérant la commande *value.split(« , »)* OpenRefine créera autant de triplets RDF (ici *skos:altLabel*) qu'il y a de valeurs de cette colonne.



Déclarez le schème

Les concepts qui seront créés doivent être attachés à un objet de type *skos:ConceptScheme*, qui représente l'ensemble de votre schème. Il faut donc leur ajouter une propriété de type *skos:inScheme* avec comme objet l'URI de votre vocabulaire (dans notre exemple <http://opendata.inrae.fr/Thesaurus/ThesaurusExemple> (cadre rouge dans la figure ci-dessous)).

Déclarez ce schème dans une nouvelle ligne du RDF skeleton en utilisant la valeur constante <http://opendata.inrae.fr/Thesaurus/ThesaurusExemple> et en indiquant type (propriété *rdf:type*) avec comme valeur *skos:ConceptScheme* (cadre vert) ainsi qu'une propriété *skos:prefLabel* pour lui donner un nom.



Déclarez les Top Concepts

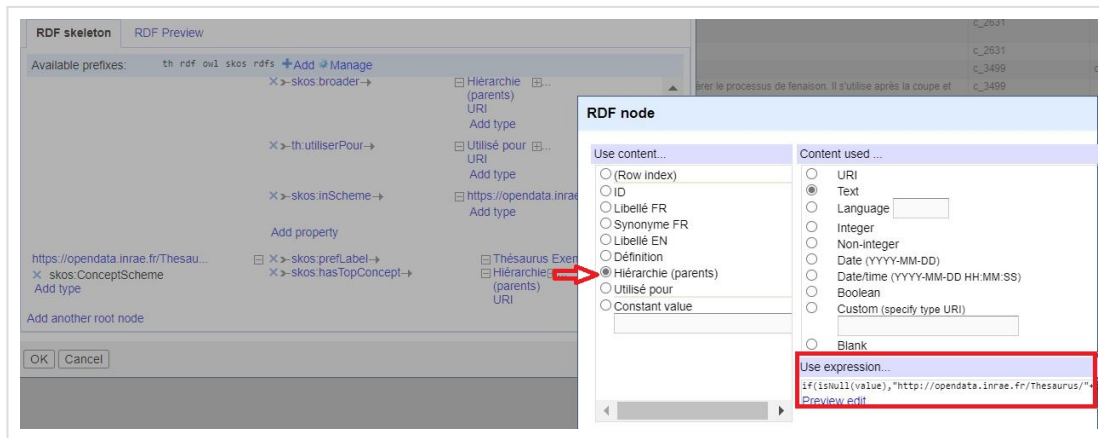
La dernière étape consiste à déclarer les concepts de 1er rang. Cette information est utile aux outils de gestion des cookies. Dans le cas d'une simple liste, tous les concepts doivent être

déclarés comme “top concept”. Dans le cas d’une hiérarchie, cela concerne uniquement les concepts qui n’ont pas de parent. Avec OpenRefine, c’est possible mais nécessite de connaître quelques fonctions.

Dans notre exemple, il faut d’abord récupérer l’identifiant du concept pour construire le triplet [URI du schéma – *skos:hasTopConcept* – URI de concept]. Pour cela, créer la propriété *skos:hasTopConcept* ayant comme sujet le *ConceptScheme*. Ensuite, pour ne sélectionner que les concepts qui n’ont pas de parent, on utilise la colonne *Hiérarchie (parents)* pour laquelle on teste l’absence de valeur avec la combinaison de fonctions suivante :

```
if(isNull(value), »http://opendata.inrae.fr/Thesaurus/ »+cells.ID.value,null)
```

et comme indiqué ci-dessous. Attention à bien sélectionner “Text” pour la partie *Content used* (car on construit l’URI manuellement).



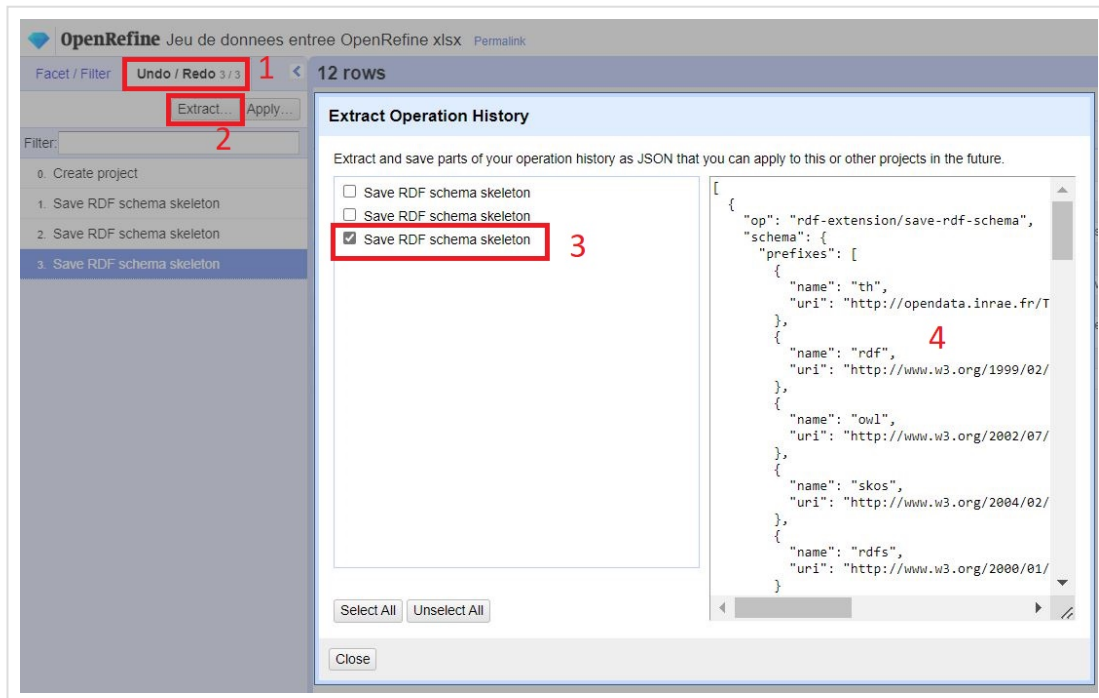
Exportez la ressource en SKOS



Vous avez fini de déclarer les étapes de transformation de votre fichier tabulé vers SKOS, félicitations. Vous pouvez maintenant exporter votre schéma représenté à l’aide du modèle de données SKOS en cliquant sur export comme indiqué ci-dessous. Vous avez alors le choix de télécharger la ressource en RDF/XML ou Turtle.

Exportez la procédure de correspondance pour la réutiliser

Il est possible de récupérer le “skeleton” qui contient toutes les déclarations et opérations associées (pour les rejouer sur un fichier similaire ou les partager) à partir de l’onglet d’historique *Undo/Redo* (1) et le bouton *Extract* (2). Sélectionnez la dernière ligne intitulée “*Save RDF schema skeleton*” (3) pour récupérer la dernière version. Recopiez le contenu de la fenêtre de droite (4) et sauvegardez-le dans un fichier texte.



Pour reproduire les étapes de notre test d’OpenRefine, vous pouvez récupérer le **code du skeleton** et l’appliquer sur le jeu de données en l’insérant dans la fenêtre accessible par le bouton “*Apply*”.

Pour aller plus loin

Liens utiles

- Télécharger OpenRefine et ses extensions : <https://openrefine.org/download.html>
- Documentation Utilisateurs d’OpenRefine : <https://docs.openrefine.org/>

Cas d’utilisation d’OpenRefine pour transformer des vocabulaires

OpenRefine a été utilisé lors du travail de transformation des données issues du dictionnaire de la Viande pour construire le Thésaurus de la Viande. Ce travail est décrit dans un article en français et un article en anglais. [Lire le billet sur le blog](#) du projet FooSIN.

Cette ressource a été développée dans le cadre du projet ANR FooSIN (ANR-19-DATA-0019-01). [A découvrir aussi : les 3 outils pour transformer un vocabulaire au format tabulé vers](#)

SKOS. Lire l'introduction à SKOS.

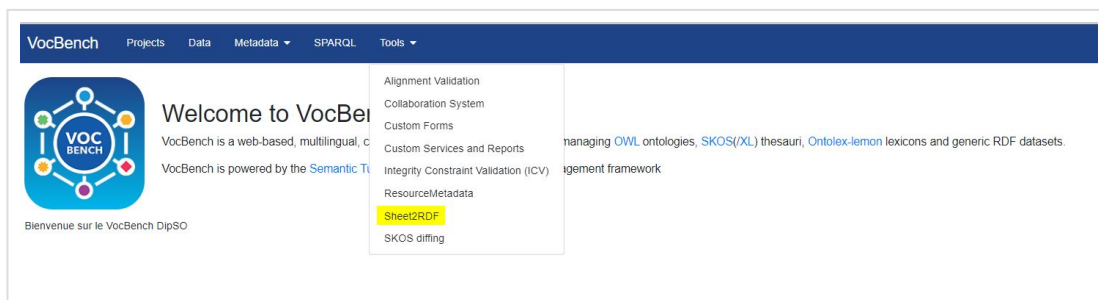


Auteurs (INRAE): Jérémy Yon, Sophie Aubin

Relecteurs (INRAE) : Catherine Roussey

Transformer un fichier tabulé en SKOS avec Sheet2RDF (VocBench)

Cette documentation « pas à pas » vous montre comment utiliser **Sheet2RDF** dans **VocBench** pour transformer un vocabulaire au format tabulé en une ressource représentée à l'aide du modèle standard SKOS. Cette page est un des résultats du [test comparatif de 3 outils pour transformer un vocabulaire au format tabulé vers SKOS](#) (projet FooSIN). Toutes les données sont publiées sur l'entrepôt ouvert recherche.data.gouv.



VocBench est un outil open source d'édition de thésaurus et d'ontologies en ligne. Le logiciel peut être installé sur serveur ou sur son poste. En plus des fonctionnalités d'édition de ressources en SKOS et OWL, VocBench embarque divers outils pour vérifier la syntaxe d'un fichier, valider l'alignement de vocabulaires ou transformer un fichier tabulé au format RDF. C'est ce dernier outil appelé **Sheet2RDF** que nous testons. La documentation officielle en anglais est consultable à [cette adresse](#).

L'objectif de ce pas à pas est de vous montrer comment produire une ressource suivant le modèle SKOS en partant d'un fichier tabulé et en utilisant le module Sheet2RDF dans l'interface VocBench (version 9.1.0). Notons que Sheet2RDF peut être utilisé en mode commande mais nous n'aborderons pas cet usage dans ce tutoriel.

Prérequis

- [Télécharger VocBench](#) et accéder à la [documentation utilisateur](#).
- [Avoir des notions de RDF et du modèle SKOS](#).

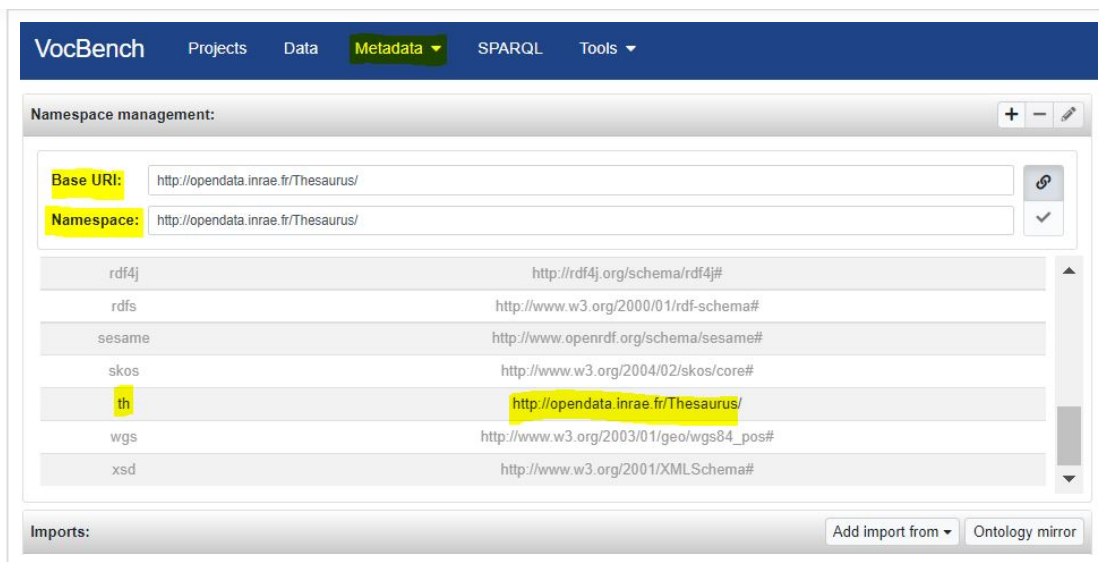
Pour réaliser la transformation, vous devez travailler au sein d'un projet VocBench paramétré pour une ressource en SKOS. Assurez-vous que l'URI de base et le nom de domaine de votre projet VocBench sont bien déclarés (onglet *metadata/Namespaces and Imports*) pour éviter les

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

ACCEPTER

REFUSER



Préparez le fichier de données

Important : chaque cellule du fichier ne doit contenir qu'une seule valeur. Si votre fichier contient des valeurs multiples, vous pouvez les séparer sur plusieurs colonnes (avec Excel ou NotePad++ par exemple) ou sur plusieurs lignes (avec OpenRefine par exemple). Dans notre exemple ci-dessous, le concept c_2631 a deux libellés alternatifs (altLabel) indiqués sur une ligne chacun, l'identifiant du concept étant répété.

	A	B	C	D	E	F	G
1	URI	skos:prefLabel@fr	skos:prefLabel@en	skos:altLabel@fr	definition	skos:broader	th:Utiliser_pour
2	c_2631	Matériel	equipment	équipement	Les objets nécessaires à une exploitation.		
3	c_2631			outillage			
4	c_25753	Matériel Agricole	farm equipment			c_2631	
5	c_25889	Broyeur	grinders		Qui permet de broyer.	c_25753	c_25675
6	c_37683	Triturateur	crushers		Appareil servant à la trituration des substances.	c_25753	c_25675
7	c_3499	Matériel de récolte	harvesters		Machine agricole qui rassemble une culture vivrière des champs	c_2631	

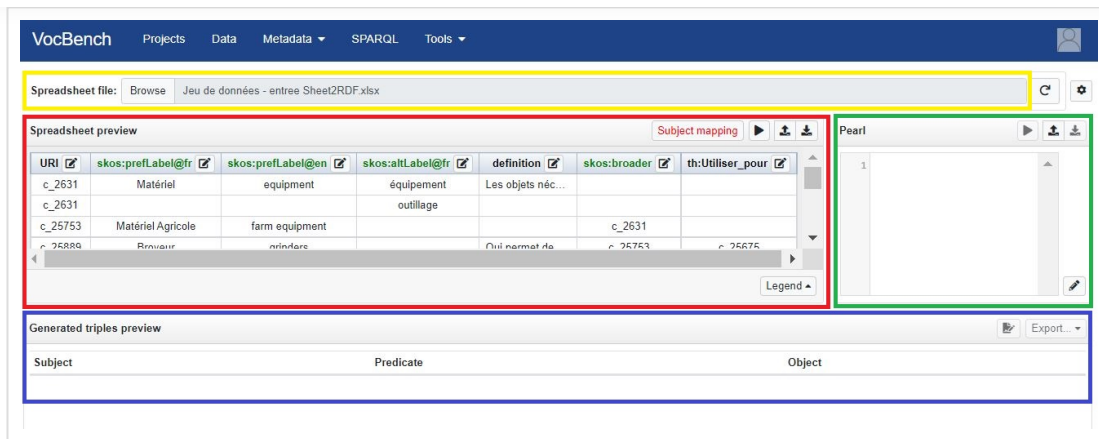
Afin de faciliter le travail de paramétrage dans Sheet2RDF, nous vous recommandons de retravailler les intitulés de colonnes du fichier source en indiquant directement les propriétés SKOS cibles ainsi que les paramètres de langue (ex : @fr), comme montré ci-dessus. De cette manière, Sheet2RDF interprètera ces informations et pré-configurera certains paramètres. Vous pouvez télécharger le fichier test contenant les données d'entrée [ici](#).

Enregistrez votre fichier au format xlsx, odt ou csv. Il doit être encodé en UTF-8.

L'interface Sheet2RDF

Accédez à VocBench depuis un navigateur. Dans le bandeau bleu, sélectionnez *Tools* puis

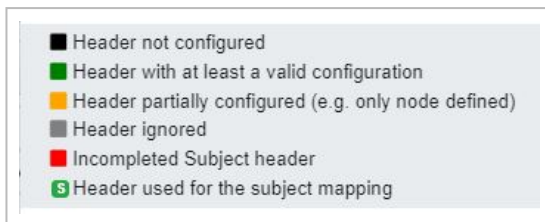
Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)



Chargez le fichier source

Pour commencer, chargez le fichier contenant votre jeu de données au format tabulé en utilisant la boîte de chargement *Spreadsheet file* (cadre jaune ci-dessus).

Dans la fenêtre *Spreadsheet preview* vous pouvez visualiser votre jeu de données et vérifier que le chargement s'est fait correctement. La configuration de la procédure de transformation se fait colonne par colonne.



La couleur de l'en-tête de chaque colonne vous permet de contrôler l'état du paramétrage : **vert** indique que la colonne est bien configurée, **rouge** et **orangé** qu'il manque des informations, **noir** que celle-ci n'est pas configurée, et **gris** que vous écarterez la colonne du traitement. La pastille **[s]** verte indique que la colonne contient le sujet des prédicats. Si vous avez suivi nos recommandations pour les intitulés des colonnes du fichier source, plusieurs en-têtes apparaissent déjà en vert.

Paramétrez Sheet2RDF

Indiquez le sujet des triplets

La première tâche est d'indiquer à l'outil quelle colonne contient le sujet des prédicats à générer. S'il s'agit d'identifiants comme dans notre exemple, ils serviront à créer les URIs des concepts. Cliquez sur le bouton *Subject mapping* qui est rouge à ce stade (car pas encore complètement configuré). Lorsque vous cliquez dessus, une nouvelle fenêtre appelée *Subject Header editor* s'ouvre

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Subject Header editor

Header:

PEARL feature:

Node ID:

Type: Assert type

Converter

Name	Description	RDF Capability	URI
DefaultConverter	Produces a resource by replacing placeholder with values passed as arguments: %s : string representation of the input. For IRIs = the string representation of the IRI, for literals it's the lexical form	node	coda.default
FormatterConverter	%n : local name in case of IRI %d : datatype IRI in case of literals %l : lang in case of language tagged literal %s : the value from the feature path	node	coda.formatter
RandomIdGenerator	Produces a resource identifiers in a random manner	uri	coda.randIdGen

Converter parameterization:

Signature

Additional predicate-object

1. Sélectionnez dans le menu déroulant de la section *Header* le nom de la colonne où se trouvent les identifiants de concepts (URI dans notre cas).
2. Cochez la case *Assert type* pour renseigner le type de ces URI.
3. Cliquez sur la puce jaune et sélectionnez le bon type dans l'arborescence : *skos:Concept* dans notre exemple.
4. La partie *Converter* permet de réaliser des transformations spécifiques pour générer les URIs. Sélectionnez *DefaultConverter* (d'autres méthodes sont possibles, voir la [documentation sur les converters](#)) et cliquez sur OK.

Le bouton *Subject mapping* devient vert et une pastille verte avec un [s] est apparue dans l'entête de la colonne contenant les identifiants. Elle signifie que cette colonne contient les sujets des triplets qui seront générés.

Cliquez sur le bouton "play" (flèche noire) de la fenêtre *Spreadsheet preview* et voyez le code généré dans la fenêtre *Pearl*. Cliquez ensuite sur le bouton "play" de la fenêtre *Pearl* et visualisez les triplets générés dans la fenêtre *Generated triples preview*. En cas d'erreur, vérifiez que le nom de domaine (namespace) est bien configuré pour votre projet.

Configurez les colonnes

Maintenant, il vous reste peut-être à configurer quelques propriétés supplémentaires pour vos concepts. Dans notre exemple, *definition* (colonne E) et *th:Utiliser_pour* (colonne G). Ces 2 intitulés ne sont pas interprétés automatiquement et nécessitent donc une configuration manuelle. Cliquez sur l'icône avec le crayon dans l'entête de colonne. Une fenêtre *Header editor* s'ouvre comme montré ci-dessous; elle ressemble à celle utilisée pour paramétrer le *subject*.

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Header editor Ignore header ⓘ

Header: definition

PEARL feature: col_4

Nodes ⓘ + -

Node ID	Converter type	Converter URI	Details

Graph applications ⓘ ✎ + -

Header contient le nom de la colonne sur laquelle vous travaillez et *Pearl feature* le nom de la variable associée qui sera repris dans le script Pearl. Le paramétrage se fait en deux étapes : dans la section *Nodes*, vous indiquerez une procédure pour convertir les données de la colonne en valeur RDF puis, dans la section *Graph Applications*, vous indiquerez comment le prédicat sera construit. Utilisez le bouton + de la ligne “Nodes” pour ajouter des règles. Elles s’affichent comme ci-dessous :

Node editor

ID:

Converter

Name	Description	RDF Capability	URI
DateConverter	Produces a typed literal (xsd:date) formatted according to the date pattern yyyy-MM-dd	literal	coda.date
DatetimeConverter	Produces a typed literal (xsd:dateTime) formatted according to the dateTime pattern yyyy-MM-ddTHH:mm:ss	literal	coda.datetime
DefaultConverter		node	coda.default
FormatterConverter	Produces a resource by replacing placeholder with values passed as arguments: %s : string representation of the input. For IRIs = the string representation of the IRI, for literals it's the lexical form %n : local name in case of IRI %d : datatype IRI in case of literals %l : lang in case of language tagged literal %s : the value from the feature path	node	coda.formatter
LangStringConverter	Produces a language tagged literal based on the converter arguments	literal	coda.langString
RandomIdGenerator	Produces a resource identifiers in a random manner	uri	coda.randIdGen
TimeConverter	Produces a typed literal (xsd:time) formatted according to the time pattern hh:mm:ss	literal	coda.time
TurtleCollectionConverter	Produces the TURTLE serialization (as an xsd:string) of the collection formed by the provided items	literal	coda.turtleCollection

Choisissez une/des procédure(s) de conversion

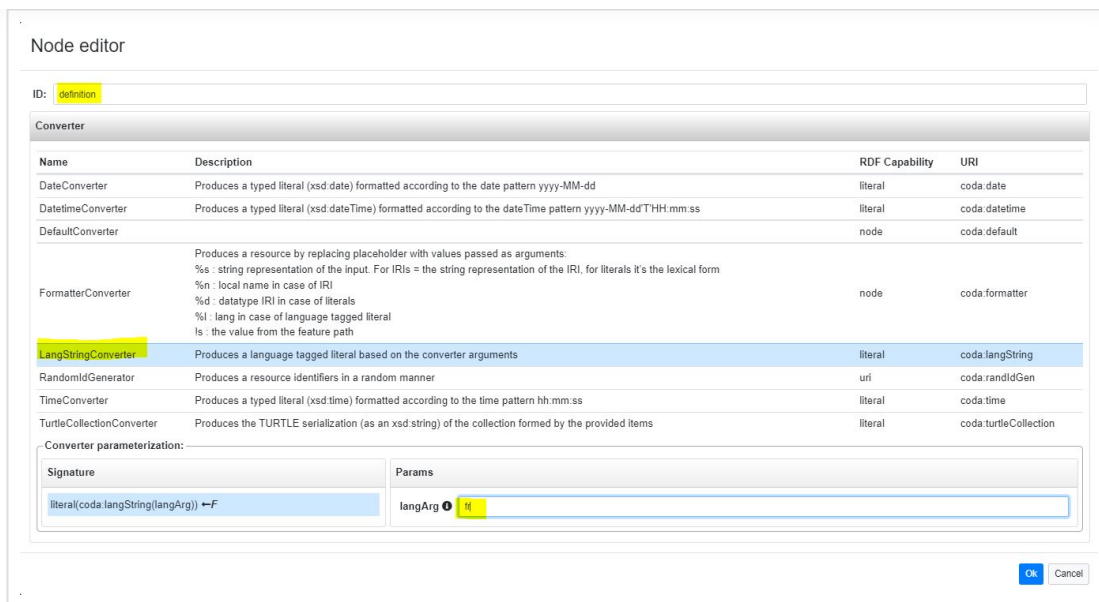
Pour chaque opération de la partie *Nodes*, indiquez un identifiant d’opération pour y faire référence ensuite et sélectionnez une procédure. Dans notre exemple, nous choisissons la procédure *LangStringConverter* qui permet d’indiquer une langue en paramètre (dans notre exemple, *fr* car toutes nos définitions sont en français – nous n’avons pas indiqué la langue dans l’en-tête comme pour les labels).

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Personnaliser mes choix

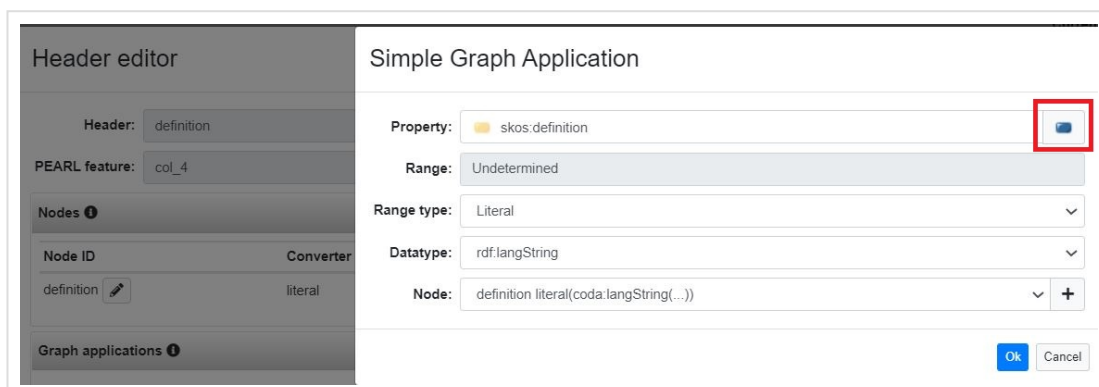
ACCEPTER

REFUSER



Déclarez le prédicat RDF

Le section *Graph applications* permet ensuite de déclarer un modèle de triplet RDF. Lorsque vous cliquez sur le “+” pour ajouter une déclaration vous avez le choix entre *graphe simple* et *graphe avancé*. Dans ce tutoriel, nous ne montrons que le *graphe simple*, suffisant dans la plupart des cas. Le *graphe avancé* donne à l'utilisateur plus de possibilités pour construire des graphes complexes, dans le cas par exemple où on souhaite générer du SKOS-XL (voir la section *Advanced graph application* dans la [documentation](#)).



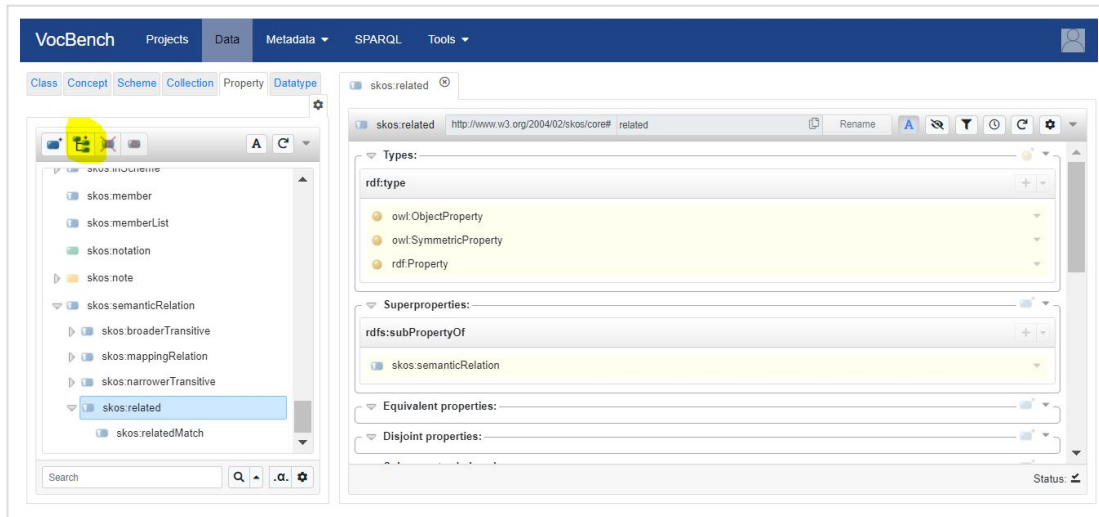
Pour paramétrer un graphe simple :

1. sélectionnez la propriété en cliquant sur le dossier bleu comme montré ci-dessous. Cette propriété est définie dans l'ontologie qui sous-tend votre projet dans VocBench. Pour la définition, nous sélectionnons la propriété *skos:definition* (déplier *skos:note*). Le champ Range est rempli automatiquement et ne peut pas être modifié.
2. Renseignez le champ *Range type* pour préciser le type de l'objet du triplet. Il ne peut prendre que deux valeurs : *Literal* (pour du texte) ou *Resource* (pour un URI). Dans notre

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Paramétrer une propriété non SKOS (th:Utiliser_pour)

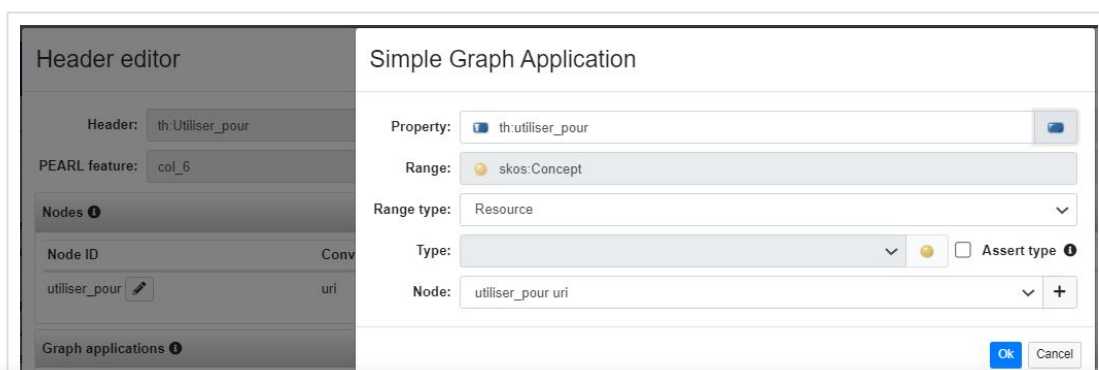
Sheet2RDF n'étant pas limité au modèle SKOS, vous pouvez déclarer des propriétés issues d'un autre modèle. Pour notre propriété *utiliser_pour*, nous choisissons de la déclarer comme une spécialisation de la propriété *skos:related*. Allez dans l'onglet "Data" du bandeau bleu et ouvrez l'onglet *Property*. Descendez jusqu'à la propriété *skos:semanticRelation* et déployez l'arbre. Cliquez sur *skos:related* puis sur le bouton *Create subProperty*. Indiquez "*utiliser_pour*" comme nom de propriété. Cliquez sur OK. Nous laisserons les valeurs par défaut de la propriété déclarée.



Retournez dans l'outil Sheet2RDF (menu Tools du bandeau bleu) et recommencez la procédure de configuration en cliquant sur l'intitulé de la colonne *th:Utiliser_pour*.

Lors du choix de la procédure de conversion (*Nodes*), sélectionnez *DefaultConverter qui* permet de transformer l'identifiant de concept en URI. N'oubliez pas d'indiquer l'ID de la procédure.

Dans la partie *Graph application*, cliquez sur + et choisissez "Simple Graph Application". Allez sélectionner *th:utiliser_pour* dans la hiérarchie des propriétés (pour rappel, sous *skos:semanticRelation/skos:related*). Pour *Range type*, choisissez *Resource*.



Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

Une fois que vous avez configuré vos colonnes – les en-têtes sont toutes vertes –, vous pouvez cliquer sur le bouton *Play* de la fenêtre *Spreadsheet preview* pour produire le script Pearl.

```
Pearl
19 prefix rdf4j: <http://rdf4j.org/schema/rdf4j#>
20
21 rule it.uniroma2.art.Sheet2RDFAnnotation id:now {
22   nodes = {
23     subject uri col_0/value .
24     col_1_node literal@fr col_1/value .
25     col_2_node literal@en col_2/value .
26     col_3_node literal@fr col_3/value .
27     definition literal(coda:langString('fr')) col_4/value .
28     col_5_node uri col_5/value .
29     used uri col_6/value .
30   }
31   graph = {
32     $subject rdf:type skos:Concept .
33     OPTIONAL { $subject skos:prefLabel $col_1_node . }
34     OPTIONAL { $subject skos:prefLabel $col_2_node . }
35     OPTIONAL { $subject skos:altLabel $col_3_node . }
36     OPTIONAL { $subject skos:definition $definition . }
37     OPTIONAL { $subject skos:broader $col_5_node . }
38     OPTIONAL { $subject <http://opendata.inrae.fr/Thesaurus/Utiliser_pour> $used . }
```

Cliquez sur le bouton *Play* de la fenêtre *Pearl* pour générer les triplets. Vous pouvez explorer les triplets dans la fenêtre *Generated triples preview*.



S'ils vous conviennent, il ne vous reste plus qu'à exporter votre ressource représentée suivant le modèle SKOS dans l'un des formats du langage RDF que vous voulez à l'aide du bouton Export (cadre rouge ci-dessus). Si vous préférez utiliser VocBench pour travailler votre ressource, le bouton Add triples (cadre vert) permet d'instancier votre base RDF d'un simple clic.

Le modèle de configuration (fenêtre *Spreadsheet preview*) ainsi que le script Pearl généré peuvent être exportés à l'aide des boutons de téléchargement et utilisés pour reproduire ce traitement à l'identique et sans effort sur un autre jeu de données similaire. Les fichiers exportés s'intitulent respectivement *s2rdf_status.json* et *pearl_export.pr*. Tous les fichiers générés pour ce pas à pas sont téléchargeables [ici](#).

Pour aller plus loin

- Télécharger VocBench : <http://vocbench.uniroma2.it/downloads/>
- Documentation Utilisateurs VocBench: <http://vocbench.uniroma2.it/doc/user/>
- Documentation Utilisateurs Sheet2RDF: <http://art.uniroma2.it/sheet2rdf/documentation/>

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)



Auteurs (INRAE): Jérémy Yon, Sophie Aubin

Relecteurs (INRAE) : Catherine Roussey

Nous utilisons des cookies sur notre site pour vous offrir l'expérience la plus pertinente en mémorisant vos préférences, les visites répétées et afin d'analyser notre trafic. En cliquant sur «Accepter», vous consentez à l'utilisation des cookies activés dans les paramètres . [Politique de cookies](#)

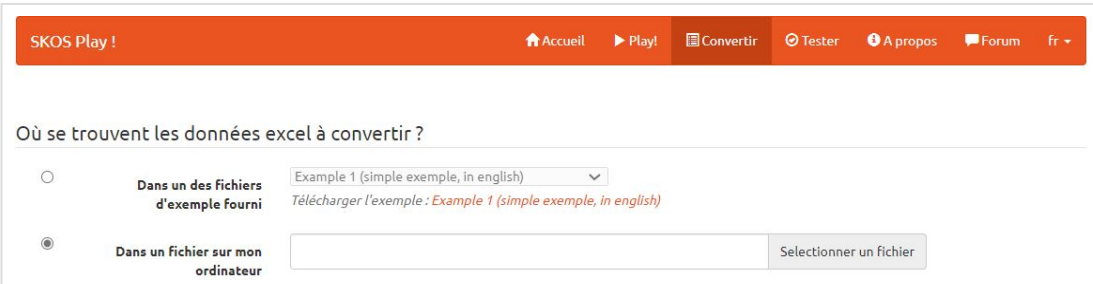
Personnaliser mes choix

ACCEPTER

REFUSER

Transformer un fichier tabulé en SKOS avec SKOS Play! Convertir

Cette documentation « pas à pas » vous montre comment utiliser **SKOS Play!** pour transformer un vocabulaire au format tabulé en une ressource représentée à l'aide du modèle standard SKOS. Cette page est un des résultats du [test comparatif de 3 outils pour transformer un vocabulaire au format tabulé vers SKOS](#) (projet FooSIN). Toutes les données sont publiées sur l'entrepôt ouvert recherche.data.gouv.



SKOS Play! Accueil Play! Convertir Tester A propos Forum fr

Où se trouvent les données excel à convertir ?

Dans un des fichiers d'exemple fourni Télécharger l'exemple : [Example 1 \(simple exemple, in english\)](#)

Dans un fichier sur mon ordinateur Selectionner un fichier

SKOS Play! est un service gratuit de visualisation de ressources représentées au moyen du modèle de données SKOS. En plus de proposer une visualisation de la ressource, SKOS PLAY vous aide à concevoir cette ressource avec des services très pratiques comme « vérifier la syntaxe du document », « vérifier le respect des règles d'intégrité du modèle SKOS » et aussi « convertir une feuille de calcul Excel en fichier SKOS ». C'est ce dernier service, disponible à l'adresse url suivante : <http://labs.sparna.fr/skos-play/convert> que nous allons tester.

Préparez le fichier de données

Pour démarrer, vous avez besoin que votre jeu de données soit enregistré dans une feuille de calcul Excel (.xls ou .xlsx) respectant une structure décrite dans la documentation SKOS Play (Excel File structure). Vous devez donc avoir préalablement choisi les propriétés RDF – SKOS ou autres – qui correspondent à chaque colonne de votre fichier (voir le tableau de correspondance pour nos données test). Ci-dessous, vous pouvez voir le début du fichier Excel qui correspond à notre jeu de données test et que nous utiliserons pour SKOS Play Convertir. La feuille de calcul Excel comprend deux parties : l'en-tête et le corps. Le fichier complet est téléchargeable [ici](#).

	A	B	C	D	E	F	G
1	ConceptScheme URI	http://opendata.inrae.fr/Thesaurus/ThesaurusExemple					
2	PREFIX	skos	http://www.w3.org/2004/02/skos/core#				
3	PREFIX	dct	http://purl.org/dc/terms/				
4	PREFIX	th	http://opendata.inrae.fr/Thesaurus/				
5	dct:autor	FooSIN					
6							
7	URI	skos:prefLabel@fr	skos:prefLabel@en	skos:altLabel@fr(separator=",")	skos:definition@fr	skos:broader	th:Utiliser_pour
8	th:c_2631	Matériel	equipment	équipement; outillage		Les objets nécessaires à une exploitation.	
9	th:c_25753	Matériel Agricole	farm equipment			th:c_2631	
10	th:c_25889	Broyeur	grinders		Qui permet de broyer.	th:c_25753	th:c_25675
11	th:c_37683	Triturateur	crushers		Appareil servant à la trituration des substances.	th:c_25753	th:c_25675
12	th:c_3499	Matériel de récolte	harvesters		Machine agricole qui rassemble une culture vivrière des champs	th:c_2631	

— Figure 1. Extrait du fichier Excel pour SKOS PLAY

Préparez l'en-tête

L'en-tête correspond à toutes lignes précédant les intitulés de colonnes et permet de déclarer :

- l'URI du schème (case B1 **obligatoire**) qui contiendra les concepts. Nous avons choisi <http://opendata.inrae.fr/Thesaurus/ThesaurusExemple> comme URI pour le schème de notre exemple.
- les préfixes des modèles dont sont issues les propriétés qui seront utilisées pour représenter les données et les métadonnées : a minima, le préfixe pour SKOS et le préfixe pour l'URI de base notre vocabulaire. Dans notre exemple, le préfixe th de la cellule B4 représente donc l'URI de base <http://opendata.inrae.fr/Thesaurus/>.
- les métadonnées (auteurs, date, licence, etc.) qui s'appliquent à la ressource dans son ensemble. Vous pouvez en indiquer autant que vous le souhaitez, une par ligne. Nous vous recommandons l'utilisation des éléments de métadonnées inclus dans le modèle [Metadata for Ontology Description and publication \(MOD\)](#).

Préparez le corps

La seconde partie du fichier, le corps, doit **impérativement** commencer par une ligne qui déclare les propriétés RDF à utiliser pour chacune des colonnes. Ces propriétés seront bien sûr les propriétés SKOS utiles mais, comme vous pouvez le voir dans la cellule G7 de notre exemple, des propriétés peuvent aussi provenir d'un autre modèle. Les propriétés seront précédées de leur préfixe suivi de ":", skos: par exemple.

La colonne A sera réservée pour la déclaration des URIs de concepts.

Gestion des langues : si vos colonnes contiennent des termes dans une unique langue, vous pouvez l'indiquer au moment de déclarer la propriété SKOS en la suffixant par un code commençant par @, par exemple pour un terme préférentiel en français, on déclarera `skos:prefLabel@fr` (voir la cellule B7 de la Figure 1).

Cellules multivaluées : pour permettre d'indiquer plusieurs valeurs au sein d'une cellule pour une même propriété, vous devez définir le séparateur utilisé à l'aide de la commande (`separator=""`). La cellule D8 de notre exemple contient 2 termes alternatifs pour le concept `th:c_2631`. Ainsi deux triplets RDF seront créés (`th:c_2631, skos:altLabel, "équipement"@fr`) et (`th:c_2631, skos:altLabel, "outillage"@fr`).

Typage des données : pour indiquer le type de données à créer (une date, un nombre,...) il est possible de l'indiquer au niveau de la propriété en la suffixant avec ^^ et un type, par exemple : ^^xsd:date (en savoir plus sur [XSD](#))

Le reste du tableau contient les concepts et leurs propriétés à transformer, i.e. vos données.

Chaque cellule de la colonne A fournit l'URI du concept qui sera le sujet de chaque triplet RDF construire. Pour les colonnes suivantes, la cellule fournit la ou les valeurs de l'objet de la propriété déclarée dans la colonne correspondante.

La valeur d'une cellule peut être :

- **un URI**, i.e. une référence à une autre ressource RDF, par exemple un autre concept comme dans la cellule F9. **Important** : pour que la valeur de la cellule soit reconnue comme un URI, elle doit commencer par "http:" (ex: http://opendata.inrae.fr/Thesaurus/c_2631) ou par un préfixe déclaré dans l'en-tête, par exemple *th:c_2631* ;
- **un littéral**, i.e. une chaîne de caractère. Par exemple, la valeur de la cellule B8 contient le littéral Matériel qui sera lié au concept identifié par l'URI *th:c_2631* par la propriété *skos:prefLabel*. Cela signifie que ce concept aura comme terme préféré "*Matériel*". Le triplet RDF (*th:c_2631, skos:prefLabel, "Matériel"@fr*) sera donc créé.

Vous avez préparé vos données, le plus difficile est fait.

Générez le fichier SKOS en RDF

1. Chargez le fichier depuis l'interface web <https://skos-play.sparna.fr/play/convert>.
2. Si vous n'aviez pas indiqué de langue dans vos données, vous pouvez en choisir une qui s'appliquera à tous vos termes, définitions et autres notes.
3. Choisissez le format de sortie à l'aide du menu déroulant
4. Cliquez sur le bouton Convertir
5. Téléchargez votre fichier

Vous pouvez vous y reprendre à plusieurs fois si nécessaire, corriger les erreurs et recharger le fichier.

Beaucoup d'options et de fonctionnalités avancées sont disponibles. Pour en savoir plus, consultez [la documentation \(en bas de la page\)](#).

Allez plus loin avec SKOS-XL

Vous pouvez accéder à des options avancées dont « générer les labels en SKOS-XL ». SKOS-XL est une extension de SKOS qui fournit un support supplémentaire pour la description des termes et leur lien aux concepts. En SKOS-XL, les termes ont eux-mêmes un URI, des propriétés, par exemple grammaticales ou historiques, ainsi que des relations vers d'autres termes ou vers des concepts. L'utilisation de SKOS-XL reste facultative.

SKOS Play Convertir propose la création d'URI de termes en s'appuyant sur l'URI du concept et un suffixe que vous définissez. Prenons comme exemple le concept avec l'URI `th:c_2631`. En utilisant SKOS-XL pour représenter le terme Matériel désignant ce concept, nous obtenons 2 triplets impliquant l'URI `c_2631-label-fr-0` généré par SKOS Play :

```
th:c_2631 skosxl:prefLabel th:c_2631-label-fr-0
th:c_2631-label-fr-0 skosxl:literalForm "Matériel"
```

Cette ressource a été développée dans le cadre du projet ANR FooSIN (ANR-19-DATA-0019-01). Accéder au [comparatif des 3 outils pour transformer un vocabulaire au format tabulé vers SKOS](#). Lire l'[introduction à SKOS](#).



Auteurs (INRAE): Jérémy Yon, Sophie Aubin

Relecteurs (INRAE) : Catherine Roussey, Sylvie Cocaud, Magalie Weber