Procédure de modification des tableaux Excel du classifieur des Interactions REMIAMES

**Date**: janvier 2024

**Auteur**: [thomas.francart@sparna.fr](mailto:thomas.francart@sparna.fr)

*Cette note explique et exemplifie comment ajouter des données dans les tableaux Excel de REMIAMES pour ajouter de nouvelles interactions.*

[Description des fichiers 1](#_Toc1)

[Principe du convertisseur Excel (« xls2rdf ») 2](#_Toc2)

[Structure générale des fichiers 2](#_Toc3)

[Nommage 2](#_Toc4)

[Entête / corps 2](#_Toc5)

[Colonnes converties / pas converties 3](#_Toc6)

[Colonnes grisées = formules Excel 3](#_Toc7)

[Expressions OWL Manchester syntax 4](#_Toc8)

[Mise en forme 4](#_Toc9)

[Cas d’usage : ajouter une nouvelle interaction 4](#_Toc10)

[Enrichissement du fichier de classes et groupes substance 5](#_Toc11)

[Correspondance des substances ANSM 5](#_Toc12)

[Enrichissement du fichier des administrations 6](#_Toc13)

[Déclaration des interactions 7](#_Toc14)

[Conversion des fichiers 8](#_Toc15)

[Test dans Protégé 8](#_Toc16)

# Description des fichiers

Les données REMIAMES sont composées des fichiers suivants :

* **classes\_et\_groupes\_substances.xlsx** contient
  + la liste des classes d’interaction
  + la liste des groupes substances
  + les regroupements de groupes substances en classes d’interaction.
* **substances\_ansm.xlsx** contient
  + La liste des identifiants de substances du Référentiel Unique d’Interopérabilité du Médicament et leur regroupement dans les groupes substance REMIAMES (référence à une URI du fichier « classes\_et\_groupes\_substances.xlsx »)
* **administrations.xlsx** contient la définition de toutes les administrations recherchées dans le cadre d’une interaction médicamenteuse. Les administrations sont définies par :
  + La classe d’interaction ou le groupe substance recherché (obligatoire) (référence à une URI du fichier « classes\_et\_groupes\_substances.xlsx »)
  + Optionnellement, la voie d’administration recherchée
  + Optionnellement, une plage min/max de dose journalière recherchée
  + Optionnellement, une plage min/max de dose maximale par prise recherchée

Ce fichier est organisé en 2 onglets. Les 2 onglets ont une structure identique, l’un est dans le cas où c’est une classe d’interaction qui est la base de l’administration recherchée, l’autre dans le cas où c’est un groupe substance.

* **thesaurus\_interactions.xlsx** contient la liste des interactions recherchées. Chaque interaction est définie par :
  + Les 2 administrations en interactions (référence à une URI du fichier « administrations.xlsx »)
  + Le contexte patient recherché (qui sera affiné au fur et à mesure que la prise en compte du contexte patient sera enrichi)
  + Les caractéristiques descriptives de l’interaction (niveau, conduite à tenir, nature du risque, etc.)

# Principe du convertisseur Excel (« xls2rdf »)

Les fichiers Excel sont convertis avec le convertisseur [xls2rdf de Sparna](https://xls2rdf.sparna.fr/rest/). L’outil donne une [documentation détaillée de la structure des fichiers Excel](https://xls2rdf.sparna.fr/rest/doc.html) pris en compte.

Les fichiers peuvent être convertis en étant uploadés dans [le formulaire de conversion en ligne](https://skos-play.sparna.fr/play/convert), toujours en cochant la case « Ignorer les post-traitements SKOS » tout en bas du formulaire :



# Structure générale des fichiers

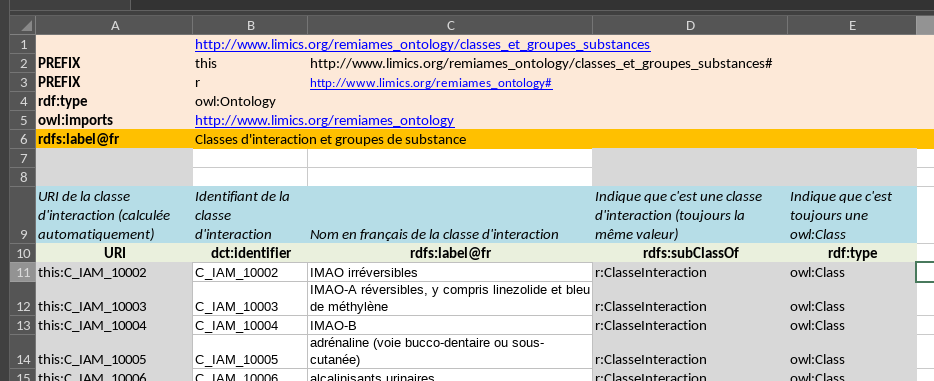
## Nommage

Il ne faut pas renommer les fichiers. Le nom des fichiers correspond à la fin de l’URI de l’ontologie correspondante décrite dans chaque fichier :

|  |  |
| --- | --- |
| **Fichier** | **Ontologie** |
| classes\_et\_groupes\_substances.xlsx | <http://www.limics.org/remiames_ontology/classes_et_groupes_substances> |
| substances\_ansm.xlsx | <http://www.limics.org/remiames_ontology/substances_ansm> |
| administrations.xlsx | <http://www.limics.org/remiames_ontology/administrations> |
| thesaurus\_interactions.xlsx | <http://www.limics.org/remiames_ontology/thesaurus_interactions> |
|  |  |

## Entête / corps

Chaque feuille de chaque fichier est composée de 2 parties : une entête et un corps. Le corps est introduit par une « ligne magique », colorée en verte, en général à la ligne 10, qui indique le mapping de chaque colonne vers un prédicat RDF au convertisseur. La « ligne magique » est précédée d’une ligne de commentaire bleue qui donne le contenu de la colonne :



L’entête déclare l’ontologie et les préfixes utilisés dans le fichier. Il n’y a pas lieu de modifier la partie entête. Pour ajouter des lignes on ajoute des lignes en fin de fichier, dans la partie « corps ».

## Colonnes converties / pas converties

Le convertisseur interprète les colonnes de la façon suivante :

* Si ce qui est indiqué dans la ligne d’entête correspond à une URI « reconnaissable » de part son préfixe (les préfixes sont déclarés dans l’entête), alors la colonne est convertie.
* Sinon, la colonne est ignorée et le convertisseur passe à la colonne suivante
* Le convertisseur s’arrête dès qu’il trouve une colonne avec une cellule vide dans la ligne d’entête.

Donc :

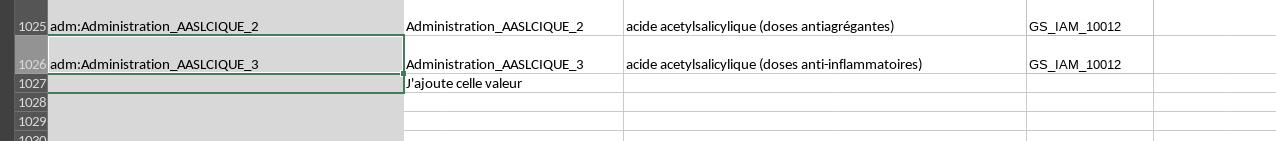
1. Il est possible d’avoir des colonnes non-converties qui servent de base à des formules qui génèrent des colonnes qui, elles, sont converties
2. Même si la colonne n’est pas convertie, on met toujours quelque chose dans la ligne d’entête, sinon les colonnes suivantes ne sont pas converties.

## Colonnes grisées = formules Excel

Les fichiers utilisent des formules Excel. Les colonnes calculées par une formule sont grisées.

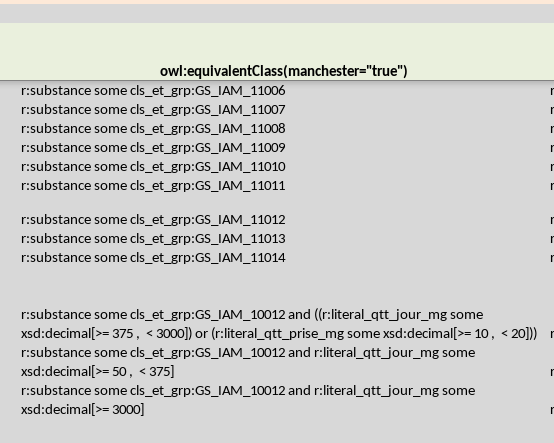
**/!\** Il faut toujours penser à descendre les formules (en « tirant » les colonnes dans Excel) des lignes grisées quand on ajoute des lignes. Ou alors copier-coller une ligne pour récupérer toutes les formules.

Par exemple si on ajoute une administration ligne 1027 dans le fichier des administration, il faut absolument « tirer » la formule de la colonne A et des autres colonnes grises dans cette ligne pour que l’URI soit calculée correctement.



## Expressions OWL Manchester syntax

Certaines colonnes contiennent des expressions OWL en « Manchester syntax », identiques à celles que l’on peut écrire dans Protégé. Ces colonnes sont marquées « ...(manchester="true") » dans la ligne d’entête.



## Mise en forme

Il n’y a aucune contrainte de mise en forme, on peut toujours colorer les cellules Excel comme on veut, et on peut figer les volets pour garder toujours la ligne d’entête visible. 2 seules petites contraintes :

* Si une ligne ou une cellule est en police ~~barrée~~, elle est ignorée par le convertisseur. Cela permet de garder des valeurs dans le tableau sans qu’elles soient converties
* Si des colonnes ou des lignes sont masquées, elles ne sont par converties. Bien penser à afficher toutes les lignes/colonnes avant conversion.

# Cas d’usage : ajouter une nouvelle interaction

Nous allons illustrer la modification des fichiers Excel par le cas d’usage suivant :

* Dans le thesaurus des interactions actuel, l’interaction « FIBRATES <-> INHIBITEURS DE L'HMG-COA RÉDUCTASE (STATINES) » est définie avec un niveau ce Contre-Indication et la note suivante : *« Contre-indication - entre le gemfibrozil et la simvastatine - pour des doses de rosuvastatine de 40 mg/ Association déconseillée - entre les autres fibrates et les autres statines - ne pas dépasser 10 mg de simvastatine (cette restriction de doses ne ... »*
* Nous allons donc découper cette interaction en 3 :
  + Une interaction de niveau « déconseillée » entre tous les fibrates et toutes les statines
  + Une première « sous-interaction » de la première, spécifiquement entre le gemfibrozil et la simvastatine, avec un niveau de contre-indication
  + Une seconde « sous-interaction » de la première, spécifiquement entre tous les fibrates et des administrations de plus de 40 mg de rosuvastatine.

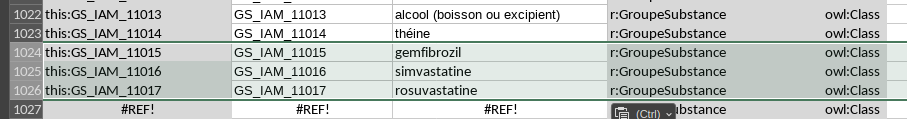
*Les fichiers Excel avec la modification sont fournis en annexe de cette doc avec les lignes additionnelles en rouge.*

## Enrichissement du fichier de classes et groupes substance

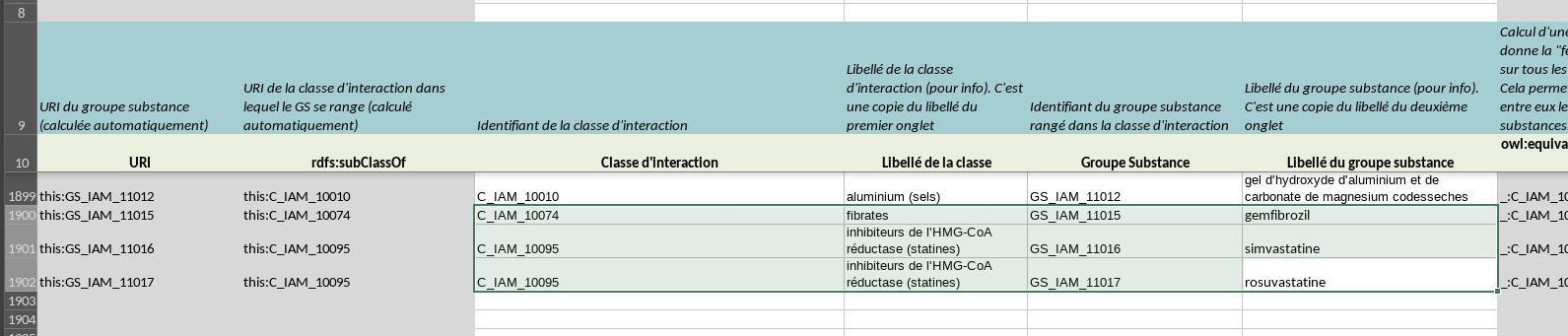
Il faut définir :

* Un groupe substance correspondant au gemfibrozil
* Un groupe substance correspondant à la simvastatine
* Un groupe substance correspondant à la rosuvastatine

On ajoute 3 lignes dans le fichier **classes\_et\_groupes\_substances.xlsx**, avec 3 nouveaux identifiants, et on s’assure que les formules des colonnes grises fonctionnent :

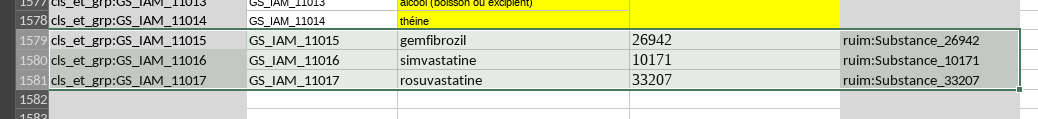


Il est également nécessaire d’indiquer que ces GS font partie d’une Classe d’Interaction (Fibrates pour gemfibrozil, statines pour les deux autres). Sans quoi les interactions que l’on créé à la dernière phase ne se « rangeront » pas correctement les unes sous les autres. Pour cela, dans le même fichier, dans l’onglet OWL-Relations, on ajoute les 3 liens entre la classe d’interaction et le GS :



## Correspondance des substances ANSM

Il faut mapper les substances ANSM vers ces nouveaux groupes en enrichissant le fichier **substances\_ansm.xlsx**. On va rechercher les codes RUIM de ces 3 substances et on les associe aux 3 identifiants de groupes substances créés précédemment :



(ici on a une correspondance 1 pour 1, en général on pourrait avoir plusieurs substances RUIM dans le même groupe substance)

(Note : a priori rien n’empecherait de déclarer une administration directement sur la base d’une URI de substance RUIM, sans passer par un groupe substance ou une classe d’interaction)

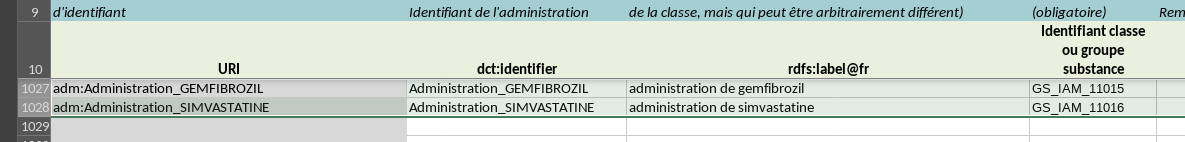
## Enrichissement du fichier des administrations

Ensuite il faut décrire les 3 administrations dont nous avons besoin pour décrire les interactions :

* Administration de gemfibrozil (sans précision)
* Administration de simvastatine (sans précision)
* Administration de rosuvastatine de plus de 40 mg

On enrichit le fichier administrions.xlsx dans l’onglet « OWL-AdminGrpSubstances » avec les 2 premières administrations « simples ».

Pour être sûr d’avoir bien toutes les colonnes calculée, on copie-colle la dernière ligne du tableau, et on ajuste les valeurs :

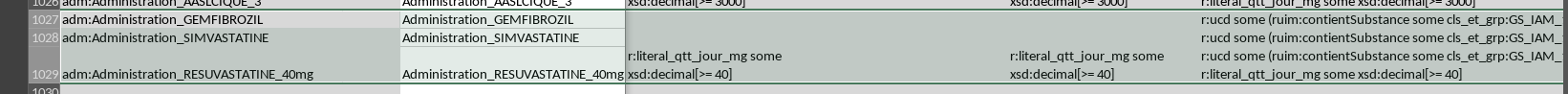


* Dans la colonne « dct :identifier » on met un identifiant pour l’administration (qui servira à son URI) – bien faire attention que cet identifiant doit être unique
* Dans la colonne « rdfs :label » on met un libellé « administration de xxx »
* Dans la colonne « identifiant classe ou groupe substance », on indique l’identifiant d’une administration créée précédemment

Pour les 40mg de rosuvastatine, on copie-colle encore une ligne, et on indique :

* « dct :identifier » = Administration\_RESUVASTATINE\_40mg
* « rdfs :label » = administration de + de 40 mg de rosuvastatine
* « identifiant classe ou groupe substance » = GS\_IAM\_11017 (l’identifiant du groupe substance créé précédemment)
* « dose journalière min en mg » = 40 (on a supposé ici que la dose indiquée était la dose journalière)

On vérifie que les colonnes grises sont bien mises à jour :



## Déclaration des interactions

Tout cela étant en place, on peut :

1. Modifier l’interaction existante pour lui donner un niveau « ASDEC »
2. La décliner en 2 « scenarios » de niveau « CI » dans les 2 cas plus précis identifiés

L’interaction existante est ligne 1053 du tableau :



1. On modifie son niveau en indiquant « asdec » dans la colonne « niveau »
2. On peut bien sûr ajuster les notes textuelles dans les colonnes « natureDuRisque » et « conduiteATenir »

On insère 2 nouvelles lignes sous celle-ci, en copiant-collant le contenu de la ligne pour récupérer toutes les formules, et on ajuste :

* Pour l’interaction « gemfibrozil – sivastatine » :
  + Scenario = « 3 » (ou tout autre identifiant permettant d’identifier ce scenario d’interaction)
  + Niveau = « ci » (c’est une contre-indication)
  + rdfs:label = « GEMFIBROZIL <-> SIVASTATINE »
  + Dct :identifier = l’identifiant de l’interaction, on peut laisser l’identifiant existant (« IAM\_10939 ») qui sera concaténé à l’identifiant du scenario pour former l’ID unique de cette interaction
  + Identifiant item A = « Administration\_GEMFIBROZIL », qui est l’ID entré dans le tableau des administrations
  + Libellé item A = « Administration de gemfibrozil » (on recopie le libellé donné précédemment)
  + Identifiant item B = « Administration\_SIMVASTATINE », qui est l’ID entré dans le tableau de administrations
  + Libellé item B = « Administration de simvastatine » (on recopie)
  + On ajuste les colonnes de notes « natureDuRisque » et « conduiteATenir »
* Pour l’interaction « fibrates – rosuvastatine + de 40mg » :
  + Scenario = « 4 » (ou tout autre identifiant permettant d’identifier ce scenario d’interaction)
  + Niveau = « ci » (c’est une contre-indication)
  + rdfs:label = « FIBRATES <-> ROSUVASTATINE + de 40 mg »
  + Dct :identifier = « IAM\_10939 »
  + Identifiant item A = « Administration\_ROSUVASTATINE\_40mg », qui est l’ID entré dans le tableau des administrations
  + Libellé item A = « administration de + de 40 mg de rosuvastatine » (on recopie le libellé donné précédemment)
  + Identifiant item B = « Administration\_C\_IAM\_10074 » (administration de fibrates)
  + Libellé item B = « fibrates » (on recopie)
  + On ajuste les colonnes de notes « natureDuRisque » et « conduiteATenir »



Note : on remarquera qu’il n’est pas nécessaire de dire que les 2 scenarios créés sont des « sous-interactions » de la première. Le raisonneur pourra le déduire de part leur définition formelle (voir la section « Test dans Protégé »).

## Conversion des fichiers

Chaque fichier doit maintenant être soumis au convertisseur à l’adresse <https://skos-play.sparna.fr/play/convert> en cochant la dernière case du formulaire « Ignorer les post-traitements SKOS ». Le résultat de chaque conversion est sauvegardé dans un fichier.

**/!\** Le convertisseur en ligne ajoute une date à la fin de chaque fichier résultat, par exemple la conversion de administrations.xlsx propose d’enregistrer un fichier administrations-2024-01-22.ttl. **Il faut supprimer la date du nom de fichier** pour qu’il s’appelle administrations.ttl, c’est obligatoire pour que le script « interactions-classifier » puisse trouver correctement les dépendances entre fichier.

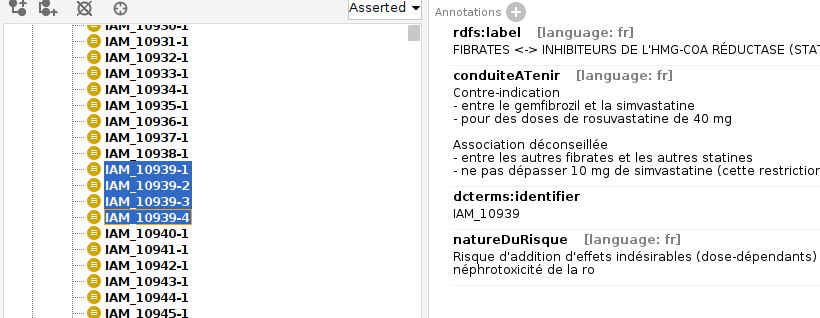
# Test dans Protégé

Pour tester dans Protégé, il faut charger tous les fichiers :

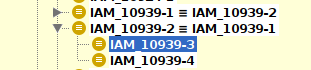
1. L’ontologie REMIAMES, **remiames\_ontology.ttl**
2. **classes\_et\_groupes\_substances.ttl**
3. **administrations.ttl**
4. **thesaurus\_interactions.ttl**

(le chargement des substances\_ansm est optionnel)

Au départ on voit les 4 interactions à plat : (note : le fichier contient déjà 2 scenarios pour cette interaction, le 1 et le 2, qui sont en fait 2 lignes identiques. Il faudrait les nettoyer)



Après avoir fait tourner le raisonneur, on voit que les scenarios 3 et 4 se rangent bien sous le 2 (qui est égal au 1) :



**Que faire si on n’a pas ce qu’on attend ?**

Il faut remonter la trace :

1. Commencer par vérifier si le GS est bien rangé sous sa classe d’interaction par le raisonneur, si l’interaction de départ était basé sur une classe d’interaction
2. Ensuite vérifier que les administrations sont bien rangées sous les administrations de départ par le raisonneur
3. Si les GS se rangent bien, les administrations vont bien se ranger, et les interactions vont bien se ranger