**Scorebot**

Documentation de la conversion en RDF

Table des matières

[Table des matières 1](#_Toc121235517)

[Prérequis 2](#_Toc121235518)

[Organisation des fichiers et répertoire 2](#_Toc121235519)

[Structure de répertoires 2](#_Toc121235520)

[Structure du répertoire « input » 2](#_Toc121235521)

[XML à convertir 2](#_Toc121235522)

[Vocabulaires contrôlés 3](#_Toc121235523)

[Structure du répertoire « work » 3](#_Toc121235524)

[Principe de la conversion 3](#_Toc121235525)

[Convertir les vocabulaires contrôlés 4](#_Toc121235526)

[Copier les fichiers 4](#_Toc121235527)

[Convertir les fichiers 4](#_Toc121235528)

[Valider les données 5](#_Toc121235529)

[Chargement des données dans GraphDB 5](#_Toc121235530)

[API Rapport Log 8](#_Toc121235531)

# Prérequis

1. Avoir [Java](https://www.java.com/fr/) installé sur le poste de travail, et avoir la commande « java » dans le PATH

# Organisation des fichiers et répertoire

## Structure de répertoires

**Dossiers**:

* input : contient les données source pour la conversion
* xslt : contient les XSLT de conversion
* shacl : description du modèle en SHACL permettant une validation des données

Deux autres dossiers seront créés lors de la conversion :

* output : contient le résultat de la conversion
* work : contient les fichiers temporaires, dont les logs, générés à chaque lancement de la conversion

Un autre dossier sera créé lors de la validation :

* output-validation : contient le rapport de validation ainsi que les logs du processus de validation

**Fichiers**:

Le répertoires contient un certain nombre de fichiers, en particuliers les scripts de conversion et de validation : scorebot\_conversion.bat et scorebot\_validation.bat

## Structure du répertoire « input »

Ce dossier est le dossier principal où se trouvent les fichiers xml à convertir.

Il contient 2 grandes familles de fichiers : les XML à convertir, en XML, et les vocabulaires contrôlés utilisés lors de la conversion, en RDF (n’importe quelle sérialisation).

### XML à convertir

Les répertoires contenant les XML à convertir sont les suivants :

* input\collectivites : doit contenir le fichier XML d’export des autorités collectivités
* input\partitions : doit contenir le fichier XML d’export des partitions
* input\personne : doit contenir le fichier XML d’export des autorités personnes
* input\thesaurus : doit contenir le fichier XML d’export du thesaurus

Chaque répertoire peut contenir 1 ou plusieurs fichiers XML. S’il y en a plusieurs, chaque fichier XML sera transformé par la XSLT correspondante.

### Vocabulaires contrôlés

Les vocabulaires contrôlés sont organisés dans les répertoires suivants :

* input\vocabulaires: contient des sous-répertoires pour chaque vocabulaire contrôlé utilisé lors de la conversion des données :
  + input\vocabulaires\iaml : vocabulaire IAML
  + input\vocabulaires\mimo : vocabulaire MIMO
  + input\vocabulaires\lexvo : vocabulaire Lexvo (langues)
  + input\vocabulaires\responsabilite : vocabulaires des responsabilités (responsabilité des musiciens)
  + input\vocabulaires\role : vocaulaire des rôles / fonctions (auteur, traducteur, etc.)
  + input\vocabulaires\educational-level : vocabulaire des niveau de compétence (premier cycle, etc.)

Chaque répertoire peut contenir 1 ou plusieurs fichiers RDF dans n’importe quelle sérialisation RDF (RDF/XML ou Turtle, etc.). Le contenu de chaque répertoire de vocabulaire contrôlé est fusionné et normalisé en un seul fichier RDF/XML avant la conversion.

## Structure du répertoire « work »

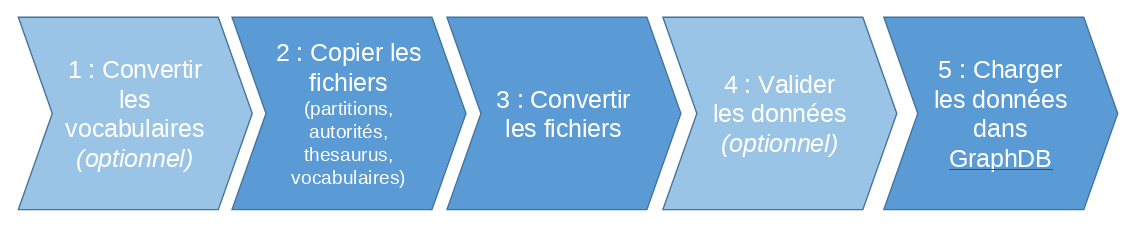
Lors du lancement du script **scorebot\_conversion.bat**, les dossiers suivants sont créés :

* work\logs : logs de la conversion des fichiers.
* work\vocabulaires\_rdf-xml : répertoire temporaires contenant les versions normalisées de tous les vocabulaires contrôlés.

Le répertoire « work » peut être supprimé sans risque.

# Principe de la conversion

Le principe général de conversion est le suivant :



# Convertir les vocabulaires contrôlés

Cette étape est optionnelle.

Si les vocabulaires contrôlés n’ont pas été modifiés, il n’est pas nécessaire de les convertir.

Pour convertir un vocabulaire contrôlé :

1. Aller à l’adresse du convertisseur <https://skos-play.sparna.fr/play/convert>
2. Charger un fichier Excel de vocabulaire dans le champ “dans un fichier sur mon ordinateur”
3. Laisser toutes les autres options par défaut et cliquer sur le bouton “Convertir” en bas du formulaire
4. Sauvegarder le fichier Turtle résultat de la conversion
5. Répéter l’opération pour les autres fichiers.

# Copier les fichiers

Copier les fichiers dans les répertoires d’input documentés ci-dessus. Chaque répertoire doit contenir (au moins) un fichier. Si certaines données n’ont pas été modifiées depuis la dernière conversion, il n’est pas utile de les mettre à jour.

# Convertir les fichiers

1. Ouvrir une fenêtre de “ligne de commandes” dans le répertoire du convertisseur.
2. Lancer **“scorebot\_conversion.bat”**
3. Vérifier que le script a généré :
   1. Des fichiers RDF résultat de la conversion dans le répertoire output.
   2. Des fichiers de log dans le répertoire work\logs
4. Vérifier les fichiers logs dans le repertoir work\logs pour vérifier s’il y a des erreurs.

# Valider les données

1. Ouvrir une fenêtre de “ligne de commandes” dans le répertoire du convertisseur.
2. Lancer **“scorebot\_validation.bat”**
3. Vérifier que le script a généré le rapport de validation dans “output-validation\scorebot\_validation-report.html”
4. Ouvrir le fichier “output-validation\scorebot\_validation-report.html” pour consulter le rapport de validation.

# Chargement des données dans GraphDB

Vérifier le fichier de validation pour s'assurer qu’il n’existe pas des erreurs. Une fois a fini de valider le fichier on doit “**uploader**” les fichiers de données à GraphDB.

A COMPLETER AVEC URL de GRAPHDB, Login et Mot de passe, et procédure pour charger chaque fichier dans un graphe nommé différent

Etapes à suivi pour uploader les fichiers :

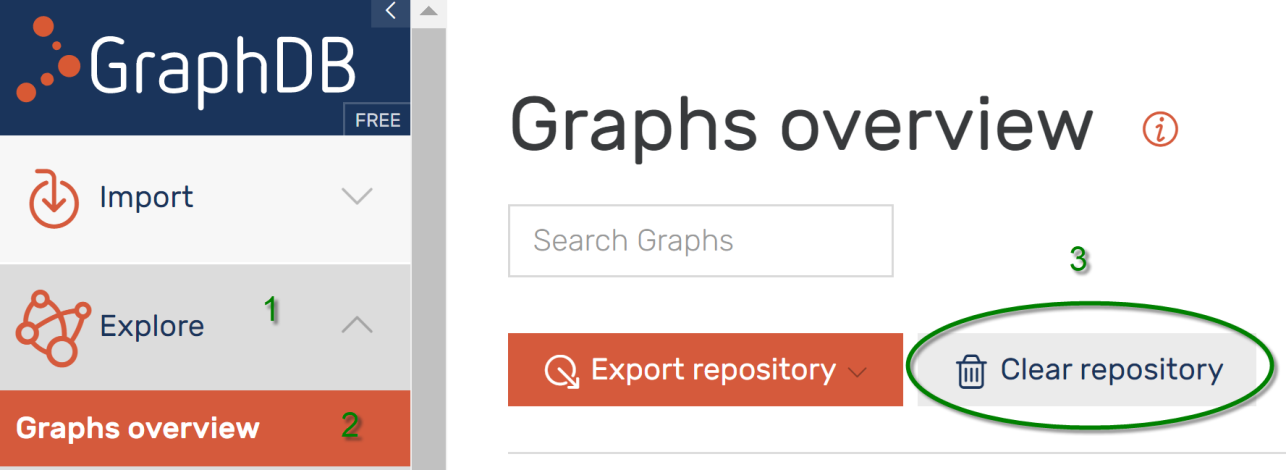
Aller à GraphDB

* <https://graphdb.prd.iumio.fr/login>
* Se connecter avec la compte d’Administrateur :
  + Utilisateur : **admin**
  + Mot de passe : **pGraphDB\_**

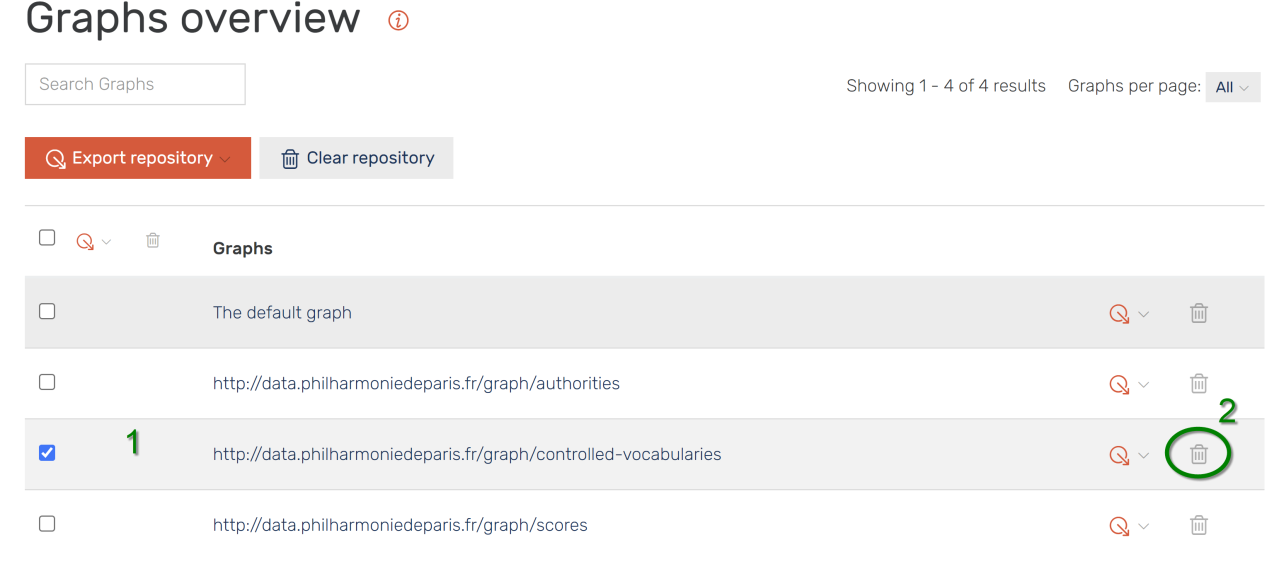
Une fois qui vous se connecte à GraphDB vous devra suivi les suivant tâches :

Valider que le repository n’a pas de données, par contraire vous devez nettoyer le repository ou graphs que vous avez besoin, avant d’uploader les fichiers :

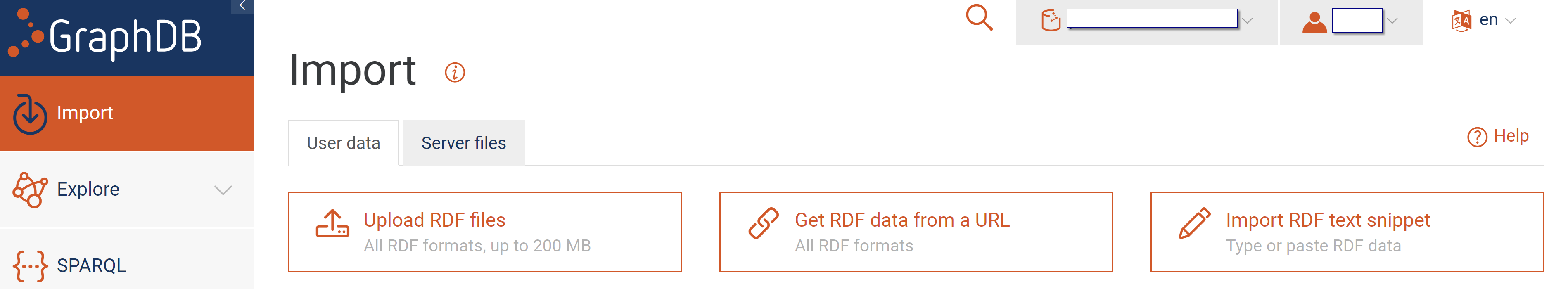
* Aller à Explorer 🡪 Graphs overview :
  + Cliquez le bouton Clear repository si vous souhaitez nettoyer le repository.

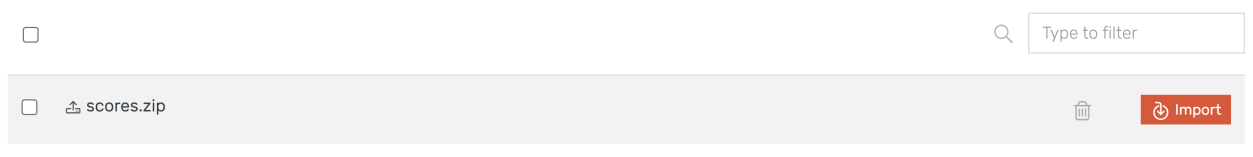


* + Si vous voulez supprimer un graph, vous devez choisir le graph (1) es cliquez sur le bouton Supprimer (2).

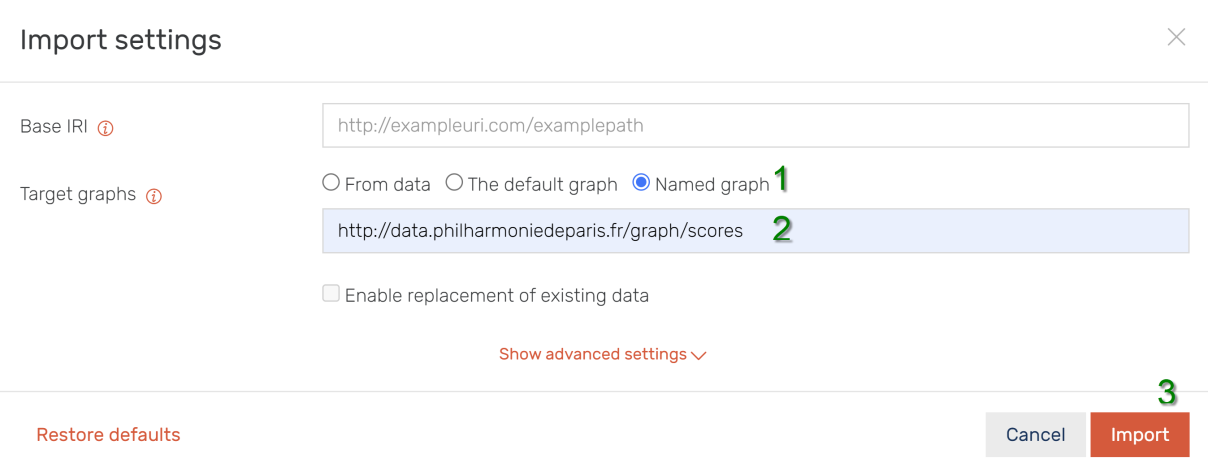


* + Confirmer l’action a demandé avec le bouton Yes.
* Uploader les données
  + Avant de commencer à uploader les fichiers, c’est important pouvoir zipper les fichiers de la suivant façon :
    - 1 fichier zip pour les fichiers concernant aux sujets des scores
    - 1 fichier zip avec tous les fichiers des autorités
    - 1 fichier zip avec tous les vocabulaires contrôlés.
  + Aller à l’option Import



* + - Cliquez sur l’option Upload RDF files et devez-vous récupérer chaque fichier zippe et uploader.
    - Uploader le fichier Score.
    - Devez-vous active la boit à gauche (1) et cliquez le bouton Import (2)



* + - Avec la fenêtre qui s’affiche, vous choisissez l’option Named graph(1), donner l’url (2), dans cette cas sera <http://data.philharmoniedeparis.fr/graph/scores> et pour terminer cliquez Import(3).
    - Pour uploader le fichier **Autorities**, devez-vous répéter le même pas précèdent (Uploader Score) et utilisant comment graphe nommé <http://data.philharmoniedeparis.fr/graph/authorities>
    - Avec le fichier de vocabulaires contrôlé, devez-vous répéter le même pas précèdent (Uploader Score) et utilisant comment graphe nommé <http://data.philharmoniedeparis.fr/graph/controlled-vocabularies>

Graphes nommés :

<http://data.philharmoniedeparis.fr/graph/scores>

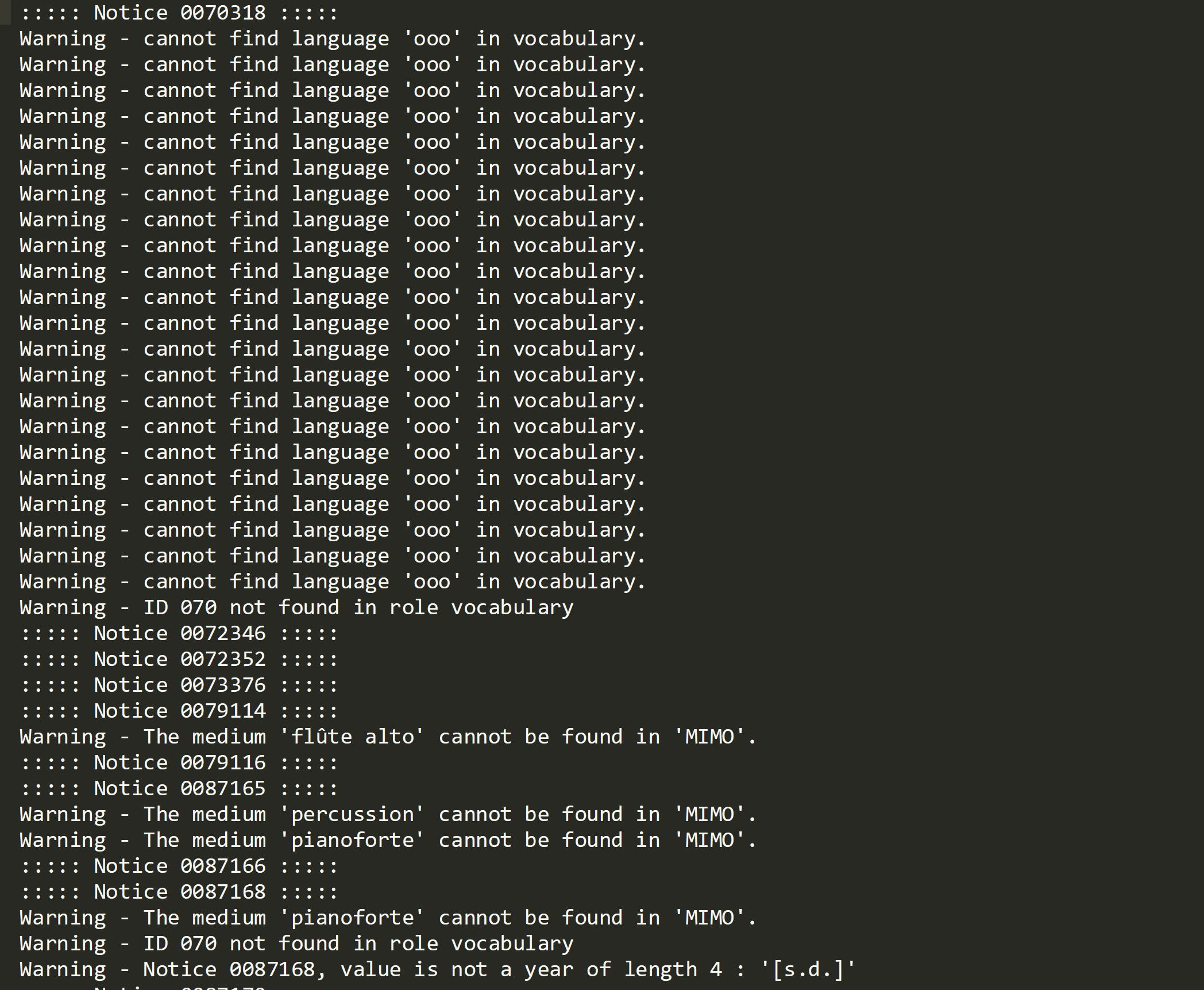
<http://data.philharmoniedeparis.fr/graph/authorities>

<http://data.philharmoniedeparis.fr/graph/controlled-vocabularies>

# API Rapport Log

L’api rapport\_log\_philhar.py est un fichier python qui servira à fournir un rapport avec tous les messages des résultats qui existe dans un fichier log.

Exemple de message:



**Outil à installer:**

Installer python <https://www.python.org/downloads/>

**Comment utiliser l’api.**

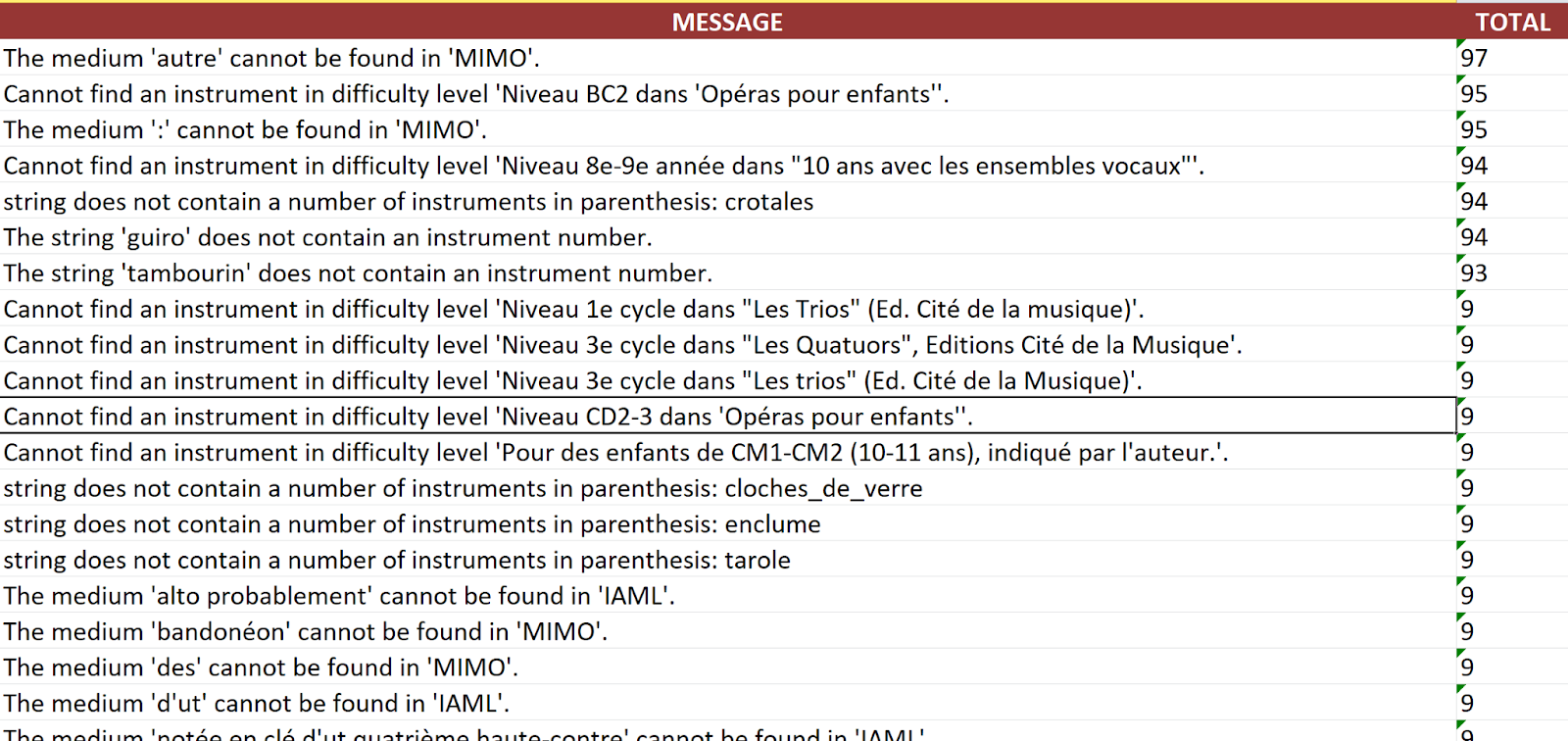
L’outil a besoin des valeurs d’entre:

* Fichier d’entre (Obligatoire): Le chemin et le nom du fichier log.
* Fichier de sortie (Obligatoire): Le chemin et le nom du fichier (avec l’extension .csv) que sauvegardera l’information.
* Optional: Si vous avez besoin de chercher un mot clé, vous pouvez l’ajouter sans problème.

Exécution avec le résultat de tous les messages trouve:

python rapport\_log\_philhar.py C:\Temp\philharmonie\log\ExportPartitions-20211201.log C:\Temp\philharmonie\scorebot.csv

Le résultat d’exécuter l’instruction précédent sera:



Exécution avec une filtre de mot clé:

python rapport\_log\_philhar.py C:\Temp\philharmonie\log\ExportPartitions-20211201.log C:\Temp\philharmonie\scorebot.csv "Les quatuors"

Le résultat:



La finalité de ce rapport est fournir une liste de messages groupée avec le total que se trouve dans le fichier log et pouvoir identifier le problème et le corrigé.