

# METADATOR



Présentation .....	2
Historique .....	2
Précisions et Précautions .....	2
Principe .....	2
Métadonnée.join([profil, shapefile]) .....	3
Manuel d'utilisation .....	4
Avant de commencer : .....	4
Récupérer le fichier zip .....	4
Lancer l'application : metadator_fr.exe .....	4
Choisir le dossier de la base de données .....	5
Choisir un profil déjà existant .....	5
Créer son profil de métadonnées pour le traitement par lot .....	6
Remplir tous les champs ou bien se baser sur un profil déjà existant .....	6
Enregistrer le profil .....	7
Choisir son profil .....	7
La patience, mère de toutes les vertus (traitement) .....	8
Le dossier ciblé et le fichier log du programme s'ouvrent : .....	8
Ouvrir les fichiers Word .....	9
Ouvrir les fichiers Excel .....	9
Importer les fichiers xml dans GéoSource .....	10
Gagner du temps avec les attributs récurrents .....	12
Sources et licence .....	13
Caractéristiques techniques .....	13
Détail des sources .....	13
Licence .....	14
Muses et références .....	14

# PRESENTATION

## HISTORIQUE

Metadator est un petit programme sans prétention, développé dans le cadre d'un projet de recherche de l'[IRD](#) afin d'automatiser la création de fiches de [métadonnées](#) de [shapefiles](#). La thématique des métadonnées est souvent aussi complexe qu'elle est essentielle. Il s'agissait donc de disposer d'un petit outil simple pour produire des catalogues de métadonnées dignes de ce nom dans un contexte (le Pérou) peu habitué à renseigner les données. Il est fortement inspiré d'un script écrit en python par [Pierre Vernier](#) qui s'appelait alors Data Inspector.

## PRECISIONS ET PRECAUTIONS

Le programme, créé à partir de 6 mois d'apprentissage de Python, n'a aucune autre prétention que de faciliter la vie du « sigiste » de tous les jours. La qualité du code est d'ailleurs certainement déplorable pour les développeurs avertis ! Il a cependant l'avantage de nous faciliter énormément le travail et d'encourager la création de métadonnées. C'est parce-que je pense qu'il peut (peut-être) dépanner d'autres personnes que je le mets à disposition.

Merci de votre indulgence qui ne doit pas interdire les commentaires et critiques constructives bien sûr.

## PRINCIPE

Metadator traite par lot les shapefiles présents dans une arborescence en en extrayant les informations techniques de base et en se basant sur des profils personnalisables pour créer des fiches de métadonnées en différents formats :

- HTML transformé en Word 2003 (.doc)
- Tableur Excel 2003 (.xls)
- Document .XML selon standard ISO19139 en partie pré-rempli

Les métadonnées en sortie sont évidemment incomplètes. Le programme n'est là que pour préparer et faciliter le travail. Les modèles de sortie ont été réalisés en fonction de nos exigences professionnelles. Le principe étant d'avoir à modifier, mettre en page ou supprimer des informations techniques (caractéristiques techniques, statistiques de base, etc.), plutôt que d'avoir à en rajouter, chercher, calculer. À l'utilisateur de juger ensuite de la pertinence de telle ou telle information. On considère que les couches d'information à traiter sont prêtes à être transmises et répondent donc à un minimum d'exigences.

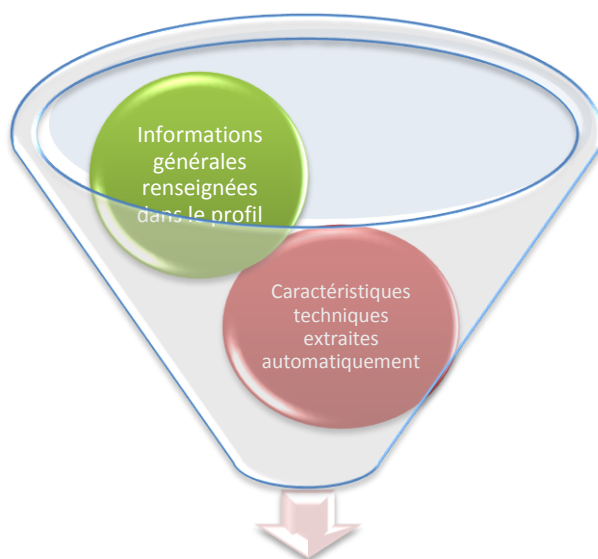


## METADONNEE.JOIN([PROFIL, SHAPEFILE])

Les métadonnées se composent généralement de deux parties :

- Les données techniques propres à chaque couche d'information : système de coordonnées, type de géométrie, attributs, etc.
- Les informations générales dépendant de l'entité responsable de la couche d'information : points de contact, historique, sources, description, etc.

Metadator part du constat qu'une bonne partie des données techniques peut être extraite et que, bien souvent, les informations générales sont redondantes, une même personne étant généralement appelée à travailler régulièrement sur les mêmes zones, thématiques, etc. D'où l'idée de procéder à partir de profils de métadonnées créables à souhait.



### Fiche de métadonnées

Lors du premier lancement, il s'agit donc de créer ses propres profils de métadonnées via le formulaire. Si le premier peut s'avérer laborieux à réaliser, il est conseillé de prendre le temps suffisant de le remplir le mieux possible. Grâce à la fonction d'import de profils, les suivants seront d'autant plus rapides à renseigner. Les informations demandées reprennent en grande partie celles requises dans l'outil [Expire](#) mais en partant du principe qu'elles sont applicables à un groupe de couches d'informations, moyennant quelques adaptations mineures. Un profil de démonstration est inclus dans le package.

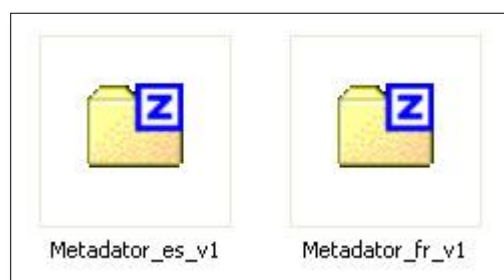
Les fiches en sortie (Word et Excel) reprennent cette organisation : une page/feuille dédiée aux informations générales, suivie d'une seconde partie détaillant les attributs et leurs spécifications techniques.

# MANUEL D'UTILISATION

## AVANT DE COMMENCER :

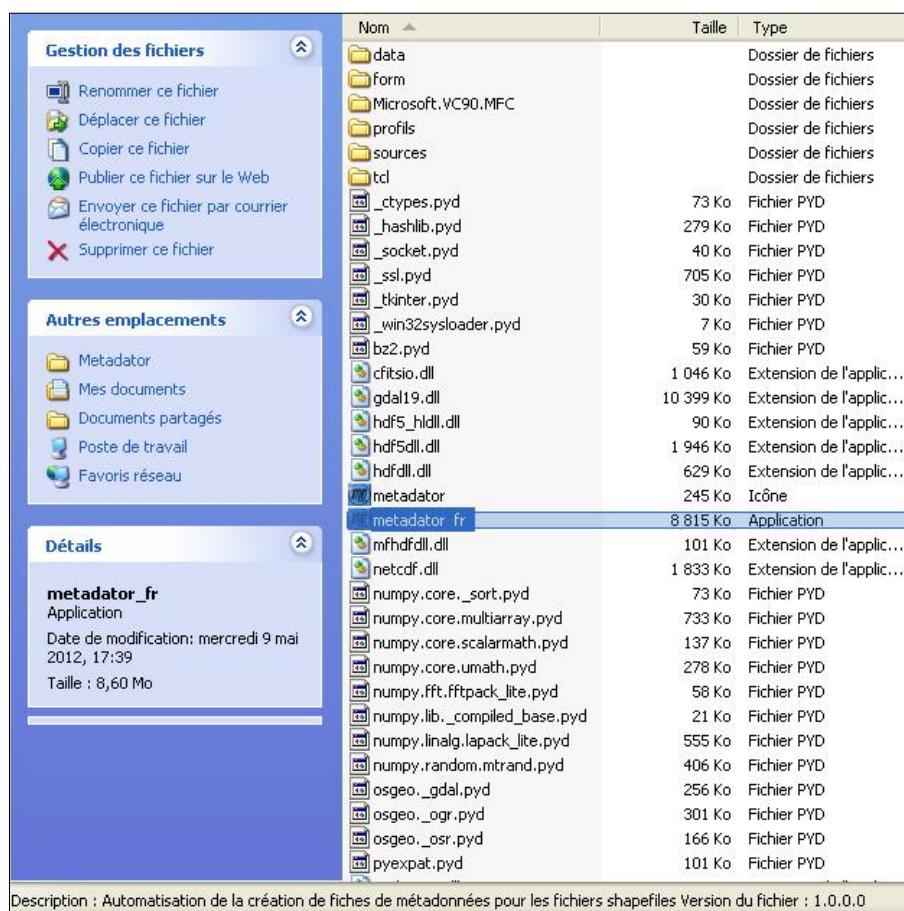
- Disposer de Microsoft Word 2003 / 2007 / 2010 (sinon [faire le nécessaire](#)) ;
- Des couches d'information géographique au format shapefiles (.shp), leurs dépendances (.shx, .dbf, ...) et projetées correctement (fichier .prj et/ou .qpj) ;
- Des noms d'attributs et de fichiers « propres » : pas d'accents, d'espaces, caractères spéciaux, etc. L'encodage c'est le mal.
- Eviter les shapes sans objet ou avec des attributs complètement vides.

## RECUPERER LE FICHIER ZIP



Développé au Pérou pour des partenaires hispanophones, l'outil existe en versions française et espagnole.

## LANCER L'APPLICATION : METADATOR\_FR.EXE



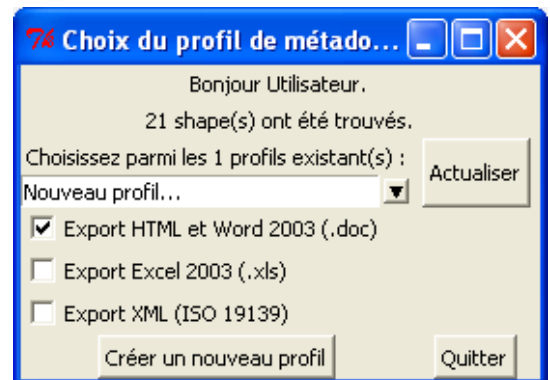
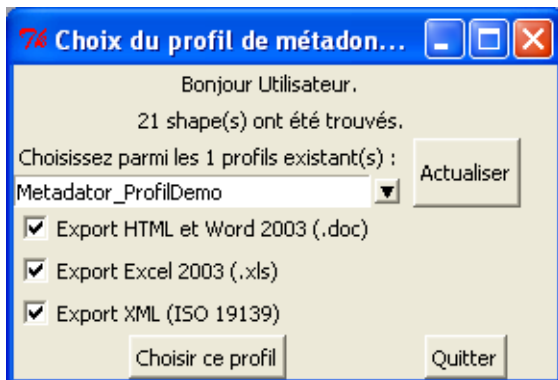
## CHOISIR LE DOSSIER DE LA BASE DE DONNEES



Metadator va alors chercher les différents shapes présents dans l'arborescence, c'est-à-dire dans tous les sous-dossiers du répertoire choisi. Attention cependant au temps de traitement qui peut rapidement s'allonger selon le nombre de shapes trouvés et leur complexité (nombre d'attributs, nombre d'objets).

## CHOISIR UN PROFIL DEJA EXISTANT

Et aller au [point 9](#) ou bien créer un nouveau profil.



Sélectionner les options de sortie désirées :

- Une fiche au format html convertie au format Word 2003
- Une fiche au format Excel 2003
- Une fiche au format xml importable dans un gestionnaire de métadonnées comme [Géosource](#).

## CREER SON PROFIL DE METADONNEES POUR LE TRAITEMENT PAR LOT

The screenshot shows the 'Création d'un nouveau profil de métadonnées' window. The 'Paramètres fichier de profil' section has a 'Nom du profil' field with the value '2012-06-05\_' and a dropdown menu set to 'Ou remplir les champs avec ce profil :'. The 'RÉSUMÉ' section has a 'Description' field with the text 'Description : description globale et commune du cadre dans lequel les données ont été créées.' and a 'Sources et historique' field with the text 'Sources et historique : origine et modalités de la création des données.'. The 'THÈMES INSPIRE' section has a list of themes on the left and an empty list on the right.

Paramètres fichier de profil

Bonjour Utilisateur.

Nom du profil : 2012-06-05\_ -- Ou remplir les champs avec ce profil : [dropdown]

RÉSUMÉ

Description : description globale et commune du cadre dans lequel les données ont été créées.

Sources et historique : origine et modalités de la création des données.

THÈMES INSPIRE

Conditions atmosphériques  
Lieux de production et sites industriels  
Régions biogéographiques  
Parcelles cadastrales  
Altitude  
Répartition de la population - démographie  
Bâtiments  
Sources d'énergie  
Ortho-imagerie  
Caractéristiques géographiques océanographiques  
Régions maritimes  
Systèmes de maillage géographique  
Caractéristiques géographiques météorologiques  
Service d'utilité publique et services publics  
Unités statistiques  
Zones à risque naturel  
Référentiels de coordonnées  
Unités administratives

## REEMPLIR TOUS LES CHAMPS OU BIEN SE BASER SUR UN PROFIL DEJA EXISTANT

The screenshot shows the 'Création d'un nouveau profil de métadonnées' window with the profile filled. The 'Paramètres fichier de profil' section has a 'Nom du profil' field with the value '2012-07-25\_ profilDemo' and a dropdown menu set to 'Metadator\_ProfilDemo'. The 'RÉSUMÉ' section has a 'Description' field with the text 'Programme PACIVUR (Programme Andin de recherche et de formation sur la Vulnérabilité et les Risques en milieu urbain) de l'IRD (Institut de Recherche pour le Développement)' and a 'Sources et historique' field with the text 'Clairefontaine ; IRD'. The 'THÈMES INSPIRE' section has a list of themes on the left and an empty list on the right.

Paramètres fichier de profil

Bonjour Utilisateur.

Nom du profil : 2012-07-25\_ profilDemo -- Ou remplir les champs avec ce profil : Metadator\_ProfilDemo

RÉSUMÉ

Description : description globale et commune du cadre dans lequel les données ont été créées.

Programme PACIVUR (Programme Andin de recherche et de formation sur la Vulnérabilité et les Risques en milieu urbain) de l'IRD (Institut de Recherche pour le Développement)

Sources et historique : origine et modalités de la création des données.

Clairefontaine ; IRD

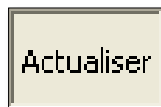
THÈMES INSPIRE

Adresses  
Altitude  
Bâtiments  
Caractéristiques géographiques météorologiques  
Caractéristiques géographiques océanographiques  
Conditions atmosphériques  
Dénominations géographiques  
Géologie  
Habitats et biotopes  
Hydrographie  
Installations agricoles et aquacoles  
Installations de suivi environnemental  
Lieux de production et sites industriels  
Occupation des terres  
Ortho-imagerie  
Parcelles cadastrales  
Ressources minérales  
Référentiels de coordonnées

Bien prendre le temps de remplir chaque champ d'information. Dans le doute, laisser la valeur par défaut, notamment pour la partie INSPIRE.

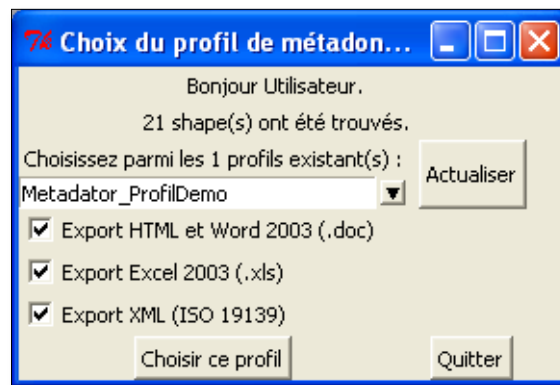
## ENREGISTRER LE PROFIL

Et « ACTUALISER » la liste des profils disponibles



## CHOISIR SON PROFIL

Puis cliquer sur « CHOISIR CE PROFIL »



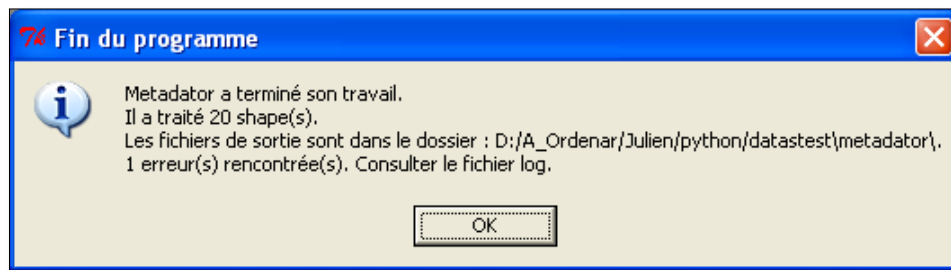
Sélectionner les options de sortie désirées :

- Une fiche au format html convertie au format Word 2003
- Une fiche au format Excel 2003
- Une fiche au format xml importable dans un gestionnaire de métadonnées comme Géosource.



## LA PATIENCE, MERE DE TOUTES LES VERTUS (TRAITEMENT)

Prendre un café ou une douche selon le nombre de couches, d'objets et d'attributs par couche à traiter. Croiser les doigts pour voir apparaître cette fenêtre :

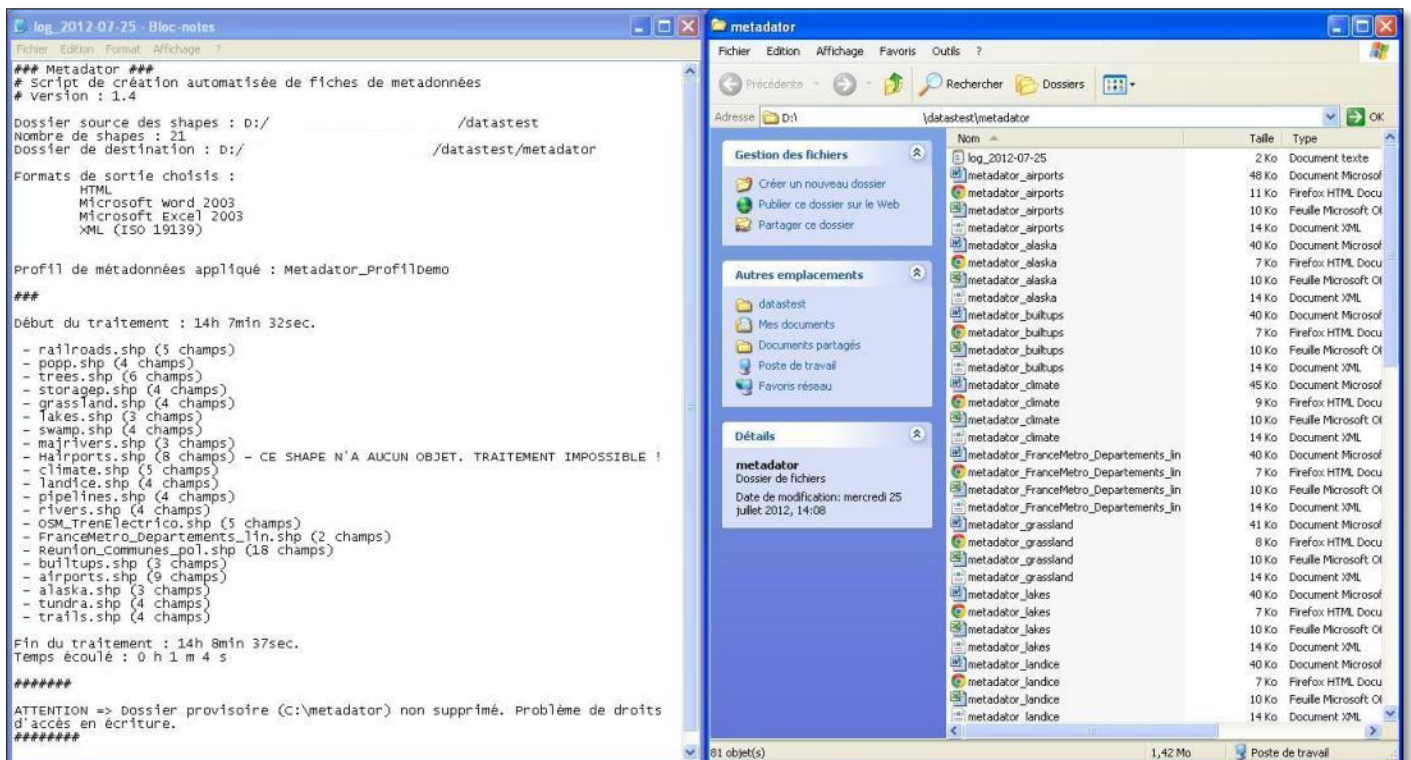


(pour le développement, les données de test utilisées ici contiennent des erreurs volontaires)

3 bons réflexes en cas de plantage :

- Les fiches sont stockées dans un dossier provisoire au format html, généralement dans C:\metadator. En consultant la dernière fiche créée on peut souvent identifier d'où vient le problème : du shape, d'un champ, etc.
- Vérifier que les données correspondent aux exigences techniques détaillées dans « [Avant de commencer](#) »
- Un fichier log d'erreur est créé automatiquement par Windows avec le code d'erreur python inscrit. Il se situe dans le dossier du programme Metadator, là où se situe l'exécutable. Le consulter et, mieux, me le faire parvenir : [julien.moura@gmail.com](mailto:julien.moura@gmail.com) !

LE DOSSIER CIBLE ET LE FICHIER LOG DU PROGRAMME S'OUVRENT :





## OUVRIR LES FICHIERS WORD

Changer le mode de lecture sur « Page », en bas à droite. Sauvegarder.



Métadonnées de Climat	
Nom du fichier	climate.shp
Mots-clé thématiques	train électrique ; métropolitain ; transports en commun
Mots-clé géographiques	Lima Callao, Pérou
Résumé	
Cadre dans lequel la donnée a été créée	Programme PACIVUR (Programme Andin de recherche et de formation sur la Vulnérabilité et les Risques en milieu urbain) de l'IRD (Institut de Recherche pour le Développement)
Nombre d'objets	14
Nombre d'attributs	5
Date de l'information	30/4/2009
Date de la dernière actualisation de la couche d'information	30/4/2009
Source(s)	IRD - SIRAD 2010 - PAVICUR 2012 - INEI - CTLC
Diffusion	interdire
Responsable de la donnée	Frère TUCK (Hadas Cie), tuck.bro@barnsdale.com
Point de contact	Robin Des BOIS (Hadas Cie), robin.hood@barnsdale.com
Site internet	<a href="#">Page du programme sur le site de PRODIG</a>
Type de géométrie	Point
Echelle	1:500000

- Pour ceux possédant un pack Office 2007 ou ultérieur, cliquer sur « **Convertir** » pour obtenir de meilleures performances.
- Remplir la partie « **Résumé** ».
- Vérifier les données et rectifier si besoin.
- Remplir les descriptions des champs manquantes.
- Ajuster les statistiques des champs : certaines sont inutiles ou non pertinentes, d'autres méritent un peu de mise en page.

## OUVRIR LES FICHIERS EXCEL

Métadonnées de Airports	
Nom du fichier	airports.shp
Mots-clés thématiques	vulnérabilité urbaine ; accessibilité ; mobilité ;
Mots-clés géographiques	Lima, Callao
Description	DESCRIPTION À COMPLÉTER
Cadre dans lequel la donnée a été créée	Programme PACIVUR (Programme Andin de recherche et de formation sur la Vulnérabilité et les Risques en milieu urbain) de l'IRD (Institut de Recherche pour le Développement)
Nombre d'objets	76
Nombre d'attributs	9
Date de l'information	18/11/2011
Date de la dernière actualisation	17/7/2012
Source(s)	Clairefontaine : IRD
Diffusion	libre
Responsable de la donnée	Frère TUCK (Hadas Cie), tuck.bro@barnsdale.re
Point de contact	Robin Des BOIS (Hadas Cie), robin.hood@barnsdale.fr
Site internet	<a href="#">Catalogue</a>
Type de géométrie	Point
Echelle	500000
Précision	40000
Système de coordonnées	Projection : Albers - Code EPSG : 4267
Extension spatiale	Min X : -4480198.52 - Max X : 4615124.98 // Min Y : 1433525.8 - Max Y : 6502586.83

- Pour ceux possédant un pack Office 2007 ou ultérieur, cliquer sur « **Convertir** » pour obtenir de meilleures performances.
- Remplir la partie « **Résumé** » de la première feuille.
- Faire un peu de mise en page.
- Remplir les descriptions des champs manquantes dans la seconde feuille et ajuster les statistiques.

## IMPORTER LES FICHIERS XML DANS GEOSOURCE

Aller dans la partie administration du catalogue. Choisir « Import en série ».

**ADMINISTRATION**

**Métadonnées & Modèle**

- Nouvelle métadonnée
- Métadonnées inutilisées
- Recherches rapides
- Ajouter une fiche de métadonnées
- Rechercher des métadonnées vides ou non utilisées
- Mes métadonnées
- Métadonnées du catalogue
- Métadonnées moissonnées
- Modèles de saisie
- Transfert de privilèges
- Transfert des privilèges de métadonnées pour un autre utilisateur
- Ajouter un standard de métadonnées
- Mettre à jour un standard de métadonnées
- Supprimer un standard de métadonnées
- Ordonner les modèles
- Ordonner les modèles de saisie
- Ajouter les modèles ou les exemples
- Sélectionner les standards :
  - csw-record
  - dublin-core
  - fgdc-std
  - iso19110
  - iso19115
  - iso19139
  - iso19139.fra
- Ajouter les modèles
- Ajouter les exemples

**Import, export & moissonnage**

- Import de métadonnées
- Import en série**
- Gestion du moissonnage
- Notification de nœuds distants
- Importer des fichiers de métadonnées (format XML ou MEF)
- Importer des fichiers XML à partir d'un répertoire local
- Ajouter/modifier/supprimer/lancer/arrêter les moissons (récupération de métadonnées depuis un nœud distant)
- Configurer la mise à jour distante de dépôt

Indiquer le répertoire cible. Si le catalogue est en ligne, il faut d'abord transférer les fiches dans un répertoire sur un serveur accessible par internet.

**IMPORTER DES FICHIERS XML À PARTIR D'UN RÉPERTOIRE LOCAL**

Répertoire: D:\datastest\metadator

Type de fichier : Fichier simple (XML, SLD, WMC...) ☒ Fichier MEF ou ZIP ☐

Actions à l'import : ☒ Aucune action à l'import  
☐ Ecraser les métadonnées de même identifiant  
☐ Générer un nouvel identifiant pour les métadonnées insérées

Feuille de style : Aucun

Valider : ☐

Affecter à ce catalogue : ☒

Groupe : Groupe exemple

Catégorie : Jeux de données

Retour Téléchargement

Si tout se passe bien, on obtient l'écran suivant :

RAPPORT SUR L'IMPORT DE MÉTADONNÉES	
	Nombre de métadonnées traitées : 20 Nombres de métadonnées insérées : 20 0 erreurs

On retrouve les fiches directement dans le catalogue.



Les champs d'information qui peuvent l'être sont pré-remplis. Il ne reste plus qu'à compléter ou rectifier.

Airports

Affichage

Actions

4696291.01700, -7357522.59330

Sud

1433525.8000

Mots clés	accessibilité (theme)
Mots clés	mobilité (theme)
Mots clés	vulnérabilité urbaine (theme)
Mots clés	Callao (place)
Mots clés	Lima (place)
Nom du système de référence	Albers (EPSG : 4267) (EPSG)

Les contacts

A l'origine de

Adresse 54 forêt de Sherwood  
Ville Nottingham  
Code postal 97418  
Pays La Réunion  
Adresse e-mail [tuck.bro@barnsdale.re](mailto:tuck.bro@barnsdale.re)

Contacts pour les métadonnées

Point de contact

Adresse 54 forêt de Sherwood  
Ville Nottingham  
Code postal 33140  
Pays France  
Adresse e-mail [robin.hood@barnsdale.fr](mailto:robin.hood@barnsdale.fr)

Les informations techniques

Contraintes d'accès et d'utilisation

Contraintes légales

Contraintes d'accès **Droit d'auteur / Droit moral (copyright)**: Restriction lié à l'exercice du droit moral  
Contraintes d'utilisation **otherRestrictions**  
Autres contraintes

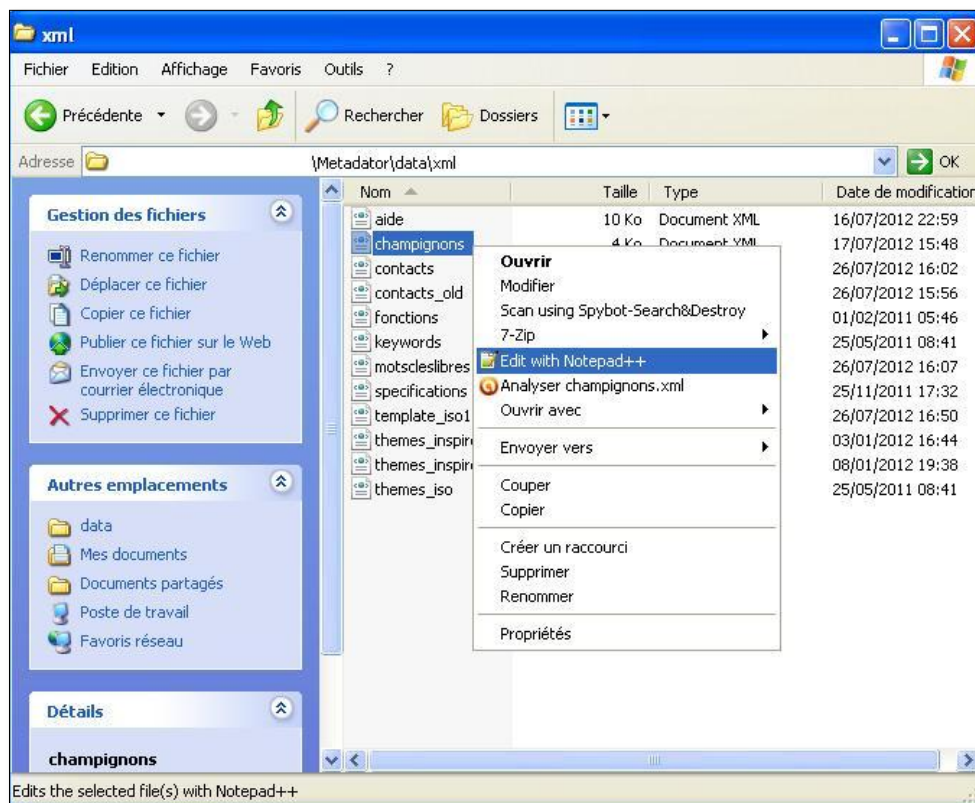
Type de représentation spatiale

**Vecteur**: Donnée vecteur (point, ligne, polygone)

## GAGNER DU TEMPS AVEC LES ATTRIBUTS RECURRENTS

Remplir les descriptions des fiches de métadonnées est déjà assez laborieux mais ça l'est encore davantage lorsque l'on a à renseigner toujours les mêmes attributs, communs à la plupart des shapes que l'on a à traiter. Metadator permet de se simplifier la vie avec la gestion des champs récurrents.

Ouvrir le fichier champignons.xml (Metadator\data\xml\champignons.xml) avec un éditeur de texte ([Notepad++](#) c'est le bien)



Structure d'un champ récurrent :

```
<champ>
  <id>01</id>
  <intitule>FID</intitule>
  <description>Identifiant universel de chaque objet.</description>
</champ>
```

Pour ajouter un champ, dupliquer cette partie et remplacer les textes entre les balises.

### Quelques recommandations :

- Respecter l'indentation de l'arborescence (retrait par rapport à la marge de la fenêtre). Rien de dramatique dans le cas contraire mais c'est beaucoup plus lisible.
- L'id (texte entre les balises <id> et </id>) doit être un chiffre entier unique sans forcément respecter l'ordre dans le document.
- Le nom du champ (texte entre les balises <intitule> et </intitule>) est sensible à la casse. Dans l'exemple ci-dessus, la description sera appliquée au champ « FID » mais pas au champ « fid ». Il faut alors dupliquer la structure en changeant avec <intitule>fid</intitule>.

# SOURCES ET LICENCE

## CARACTERISTIQUES TECHNIQUES

Metadator est développé en langage Python 2.7.x (<http://www.python.org/download/releases/2.7.3/>) avec l'aide de PyScripter (<https://code.google.com/p/pyscripter/>).



Il est distribué sous forme d'exécutable Windows sans installateur préalable. Les sources étant incluses, il est donc facile pour un utilisateur Linux, Mac ou même Windows de l'adapter à son système et sa configuration.

Pour Linux par exemple :

- désactiver systématiquement la fonction d'export au format Word en commentant les lignes de code correspondantes (# au début des lignes) :

```
Ligne 51.      from win32com.client import Dispatch      # pour la conversion au format word
```

```
# si option choisie conversion du fichier html en word
Ligne 1536.  if opt_doc.get() == 1:
                conv_word(prov +
                           "\metadator_" +
                           dico_infos_couche['nom'][:-4] +
                           ".html",
                           dest)
```

- rediriger la fonction du subprocess qui déclenche l'ouverture du formulaire sur le script correspondant dans la fonction **choix\_profil()** :

```
Ligne 151.      subprocess.Popen([chemin_script + r'\form\form_fr.exe'])
```

## DETAIL DES SOURCES

Metadator se base sur les librairies intégrées suivantes :

- tkinter et associés pour les interfaces graphiques (GUI pour les intimes) :
- os, glob, sys et shutil pour la gestion des répertoires et fichiers systèmes
- subprocess pour les sous-exécutions
- datetime et time pour les paramètres temporels
- elementtree pour les fichiers XML
- operator pour mieux gérer les itérations dans les dictionnaires
- copy pour des possibilités de copie avancée des objets
- codecs pour les (éternels) problèmes d'encodage des fichiers

Il s'appuie également sur de nombreux modules et librairies externes qui enrichissent le langage :

- [gdal / ogr / osr](#) , le couteau suisse open-source pour gérer les données géographiques :
- [dbfpy](#), pour tout ce qui a trait aux fichiers .dbf
- [pmw](#), pour étendre les possibilités de tkinter
- [dateutil](#), pour étendre les possibilités de datetime et time
- [win32client](#), pour motoriser les applications windows, notamment du pack Office
- [xlwt](#), pour écrire des fichiers Excel 2003
- [numpy](#), pour les opérations mathématiques
- [py2exe](#), pour la création de l'exécutable

Pour les utilisateurs de Windows, un site compile des installeurs exécutables pour la plupart des modules, pour les différentes versions du système d'exploitation : <http://www.lfd.uci.edu/~gohlke/pythonlibs/>

## LICENCE

Certaines librairies (dll) compilées dans l'exécutable sont sous le coup d'une licence propriétaire, notamment celles permettant de motoriser le pack Office. Ne connaissant rien à toutes ces histoires de licence informatique, je suis incapable actuellement de rentrer dans les détails. Cependant, le code python est libre et les volontaires sont cordialement invités à améliorer Metadator !

## MUSES ET REFERENCES

L'incontournable livre de Gérard Swinnen pour apprendre le Python : <http://inforef.be/swi/python.htm>

GeoTribu : <http://www.geotribu.net/>

PortailSIG : <http://www.portalsig.org/>

Expire : <http://sourceforge.net/projects/expire/>

Sam&Max un site au ton décalé plein de ressources utiles : <http://sametmax.com/>