

# Qualité des données à l'OSU OREME

Vers des données compréhensibles,  
interopérables et enrichies

# Données de qualité ?

fiables

**compréhensibles**

qualifiées

sauvegardées

**interopérables**

intègres

requêtables

**ré-utilisables**

accessibles

# Principe

- Décrire finement les données
  - Qui, pourquoi, où, comment, quoi ?
  - Stockage dans les bases de données (tables spécifiques)

Compréhension et ré-utilisabilité des données

- Utiliser des standards
  - Référentiels, voc. contrôlés, thésaurus, mots-clés, ontologies

Interopérabilité sémantique

Enrichissement, déduction d'information

**=> Rechercher, croiser, comparer**

# Descripteurs et référentiels

Métadonnées  
"bas niveau"

Thématiques

Thèmes Inspire

Mots-clés

Gemet

Agrovoc

Envthes

Stations, géographie

Coordonnées

BD Geofla  
(adm)

BD Carthage  
(hydro)

Personnes,  
organisations

Instruments

GCMD (NASA)

Paramètres,  
protocoles

Essential Biodiversity  
Variables

GCMD  
(NASA)

Ontologies /  
thésaurus métier

Unités

Unit SI

Unit Ontology

Spécifiques : taxon,  
gènes ..

Catalogue of Life

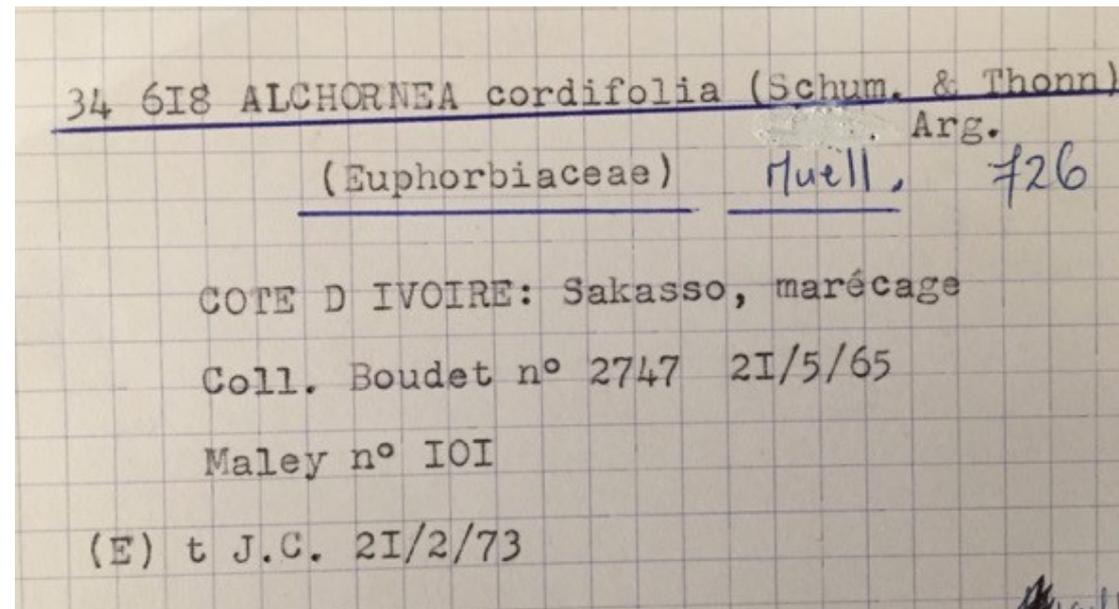
Taxref

NCBI

# Enrichissement des données : Ex 1

## Standardisation de taxons

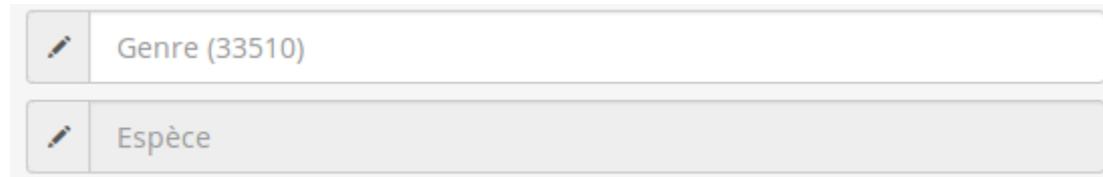
- Contexte : collection de lames de pollens et photos des grains
- Référentiel **Catalogue of Life**
  - Identifiant
  - Classification taxonomique
  - Statut, auteurs, etc ...



# Enrichissement des données : Ex 1

## Standardisation de taxons

- Cas d'utilisation 1 : **correction** orthographique des données existantes et **suggestion** de taxons proches
  - Script R, fonction *agrep* (similitude entre chaînes de caractères)
- Cas d'utilisation 2 : aide à la saisie, autocomplétion, **vérification**



The image shows a screenshot of a data entry interface. It consists of two horizontal input fields stacked vertically. The top field is labeled 'Genre (33510)' and contains a small pencil icon on the left side. The bottom field is labeled 'Espèce' and also contains a small pencil icon on the left side. The fields are light gray with rounded corners and a thin border.

# Enrichissement des données : Ex 1

## Standardisation de taxons

- Cas d'utilisation 3 : recherche de photos au niveau d'une famille, ..

Recherche de taxons

**Famille**  
  
Tapez les premières lettres de la famille recherchée (ex : Oleaceae) puis sélectionnez-la dans la liste

**Genre**  
  
Tapez les premières lettres du genre recherché (ex : Olea) puis sélectionnez-le dans la liste

**Taxons** 15 Taxon(s)



©Karl Henga / ISEM  
Alternanthera echinata



©ISEM 1 / ISEM  
Alternanthera ficoidea

### Galerie du taxon *Alternanthera echinata*

#### Taxonomie ([Catalogue of Life](#))

<b>Classification</b>	Tracheophyta > Magnoliopsida > Caryophyllales > Amaranthaceae > Alternanthera
<b>Auteur</b>	Sm.
<b>Statut</b>	Accepted name
<b>Source</b>	ChenoBase

Donnée au rang espèce => recherche / analyse à un rang supérieur

# Enrichissement des données : Ex 2

## Standardisation de cours d'eau

- Contexte : échantillonnage de truites dans des cours d'eau pour analyses génétiques
- Référentiel **Carthage** (réseau hydrographique IGN) :
  - Identifiant pour chaque entité hydrographique
  - Fiche descriptive du cours d'eau (Sandre) : `<base d'url> + <identifiant>`
  - Arbre confluence / affluence d'un cours d'eau (Sandre)



# Enrichissement des données : Ex 2

## Standardisation de cours d'eau

- Cas d'utilisation 1 : déduction automatique de l'arbre de confluence / affluence d'un cours d'eau
  - Trigger + fonction plsh qui appelle un web service du Sandre

Avant :  
Station x est sur le Coudoulous

Après :  
Station x est sur le [Coudoulous](#)  
affluent de [L'Arre](#) > [L'Hérault](#)

# Enrichissement des données : Ex 2

## Standardisation de cours d'eau

- Cas d'utilisation 2 : déduction des entités hydro les + proches d'un point : outil cartographique + fonction PostGIS

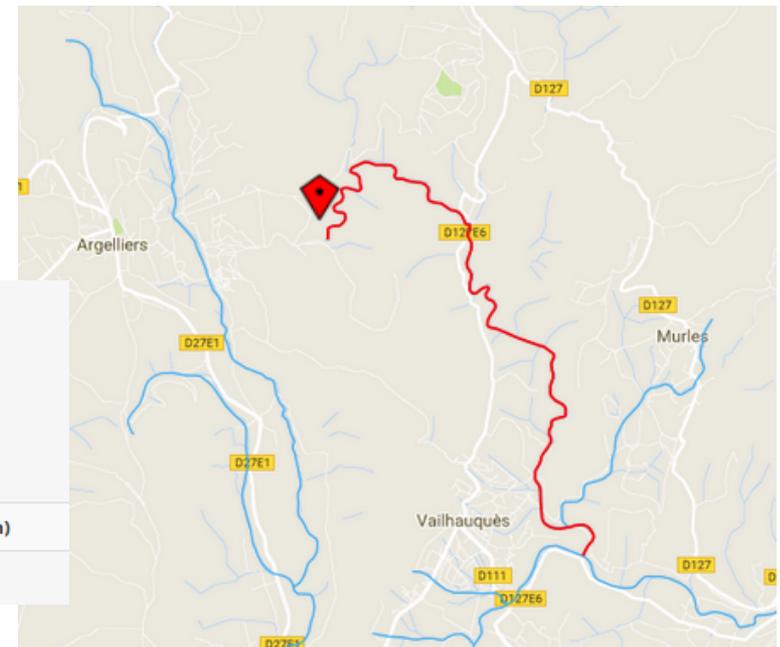
Avant :  
Station x (X, Y) est sur ?

Après :  
Station x (X, Y) est sur le [Coudoulous](#)  
affluent de [L'Arre](#) > [L'Hérault](#)

X  Y  Rayon (en m)

1 entité(s) hydrographique(s) trouvée(s) dans un rayon de 200 m

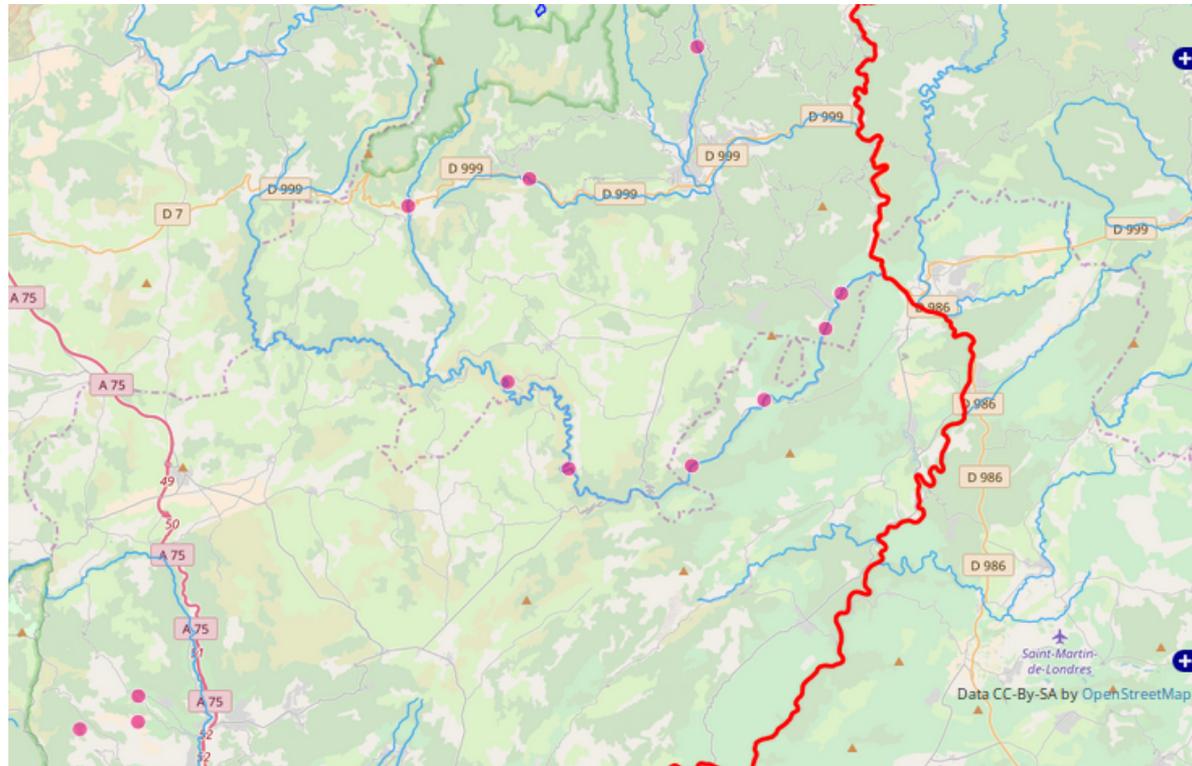
NAME	CODE	TYPE	DISTANCE (m)
L'Arnède	Y3100520	river	117



# Enrichissement des données : Ex 2

## Standardisation de cours d'eau

- Cas d'utilisation 3 : Recherche de données au niveau d'un bassin versant



Donnée au niveau entité hydro => analyse/comparaison à un niveau supérieur

# Conclusion

## Descriptions fines standardisées

- Améliorent la **qualité** des données (données fiables, non ambiguës)
- Assurent **l'interopérabilité** des données
- Permettent de **déduire** de nouvelles informations et de raisonner à des niveaux d'échelle différents
- Aident à la **saisie**
- Bonus : permettent de générer automatiquement des métadonnées standardisées
  - Ex : fiches ISO 19115 et Inspire dans un catalogue Geonetwork (bibliothèques R geometa et geonapi - Emmanuel Blondel)