InteRessources : une plateforme de catalogage de ressources linguistiques

Sarra El Ayari, Clément Plancq

InteRessources : une plateforme de catalogage de ressources linguistiques

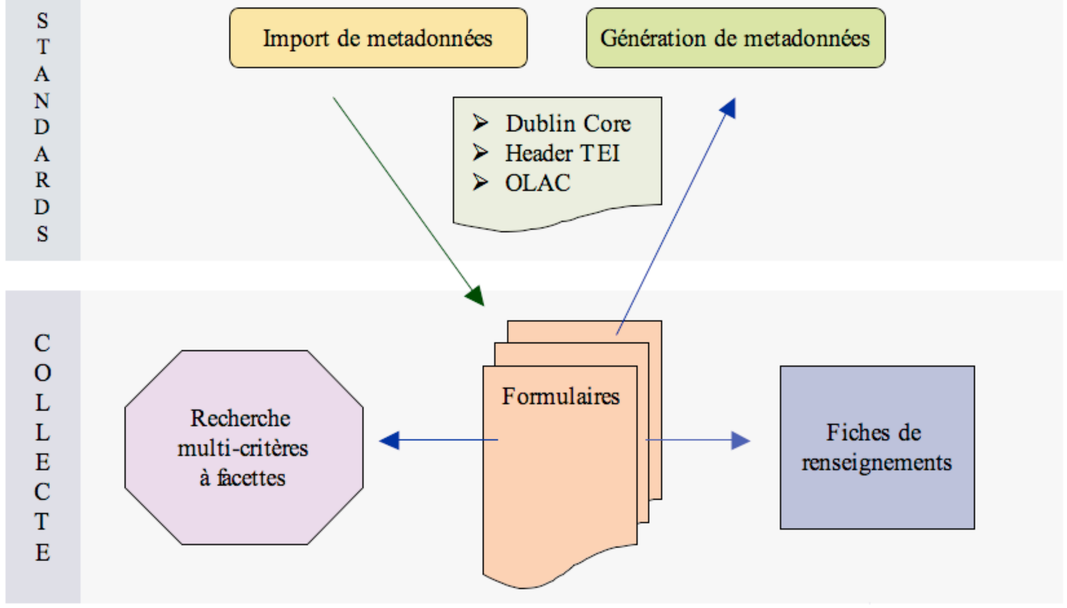
Sarra El Ayari (SFL)& Clément Plancq (LATTICE)

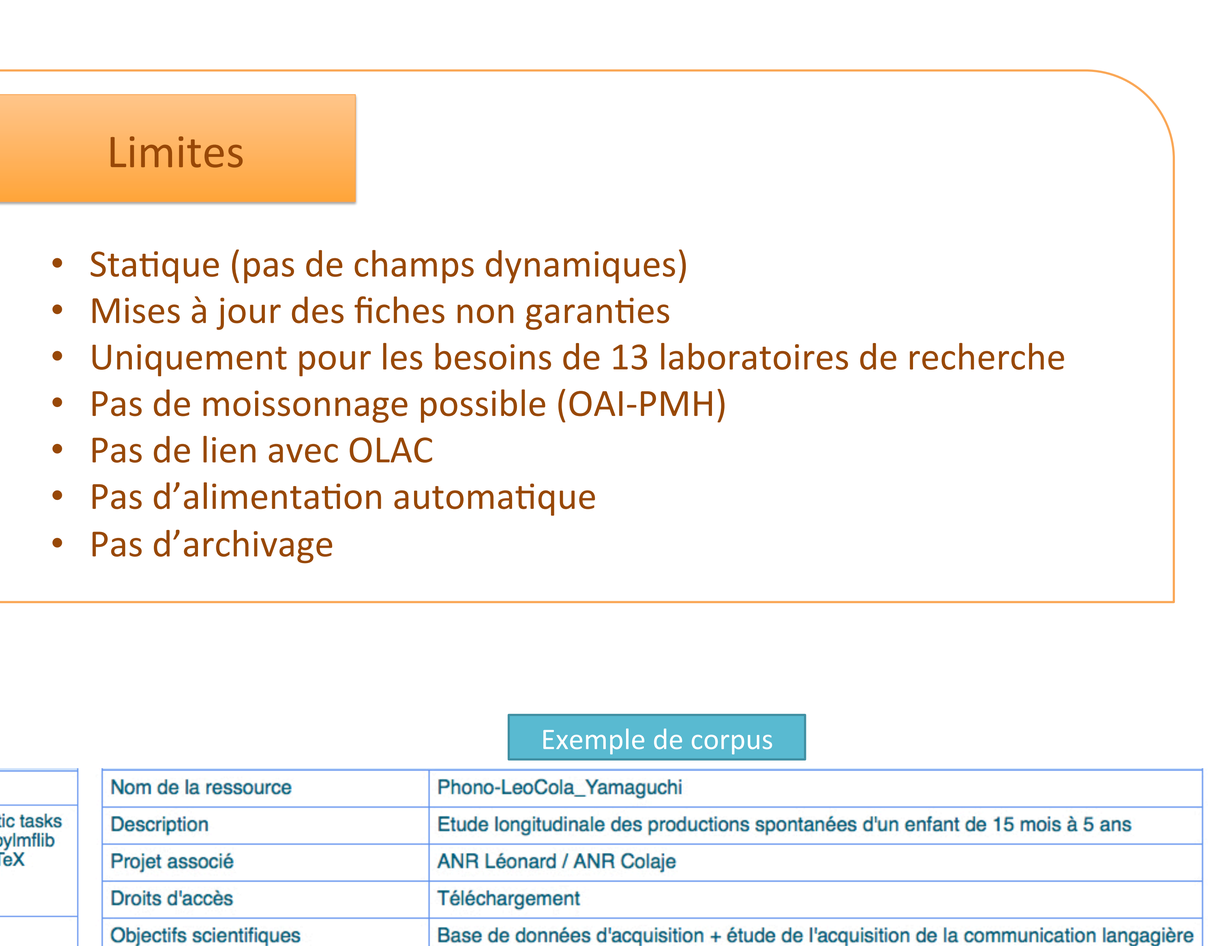
Contexte

- LabEx Empirical Foundations of Linguistics (EFL)
- Axe transversal sur les ressources linguistiques
- 13 laboratoires partenaires
- Catalogage des ressources utilisées localement : données + outils
- 90 ressources recensées :
  - 45 corpus, 30 outils, 11 lexiques, 4 dictionnaires

Objectifs

- Recensement de l'existant
- Identifier des personnes ressource
- Interopérabilité des métadonnées
- Utilisation des standards de métadonnées :
  - DC, OLAC, TEI Header, CMDI





Fonctionnalités

- Interface en ligne + base de données (HTML5 / CSS / PHP / SQL)
- **Edition** par formulaires web avec descripteurs simples
- Saisie rapide des informations
- **Recherche** à facettes
- URL stables
- **Export** automatique des métadonnées (DC, OLAC, TEI Header, CMDI) via XML et XSLT

Limites

- Statique (pas de champs dynamiques)
- Mises à jour des fiches non garanties
- Uniquement pour les besoins de 13 laboratoires de recherche
- Pas de moissonnage possible (OAI-PMH)
- Pas de lien avec OLAC
- Pas d'alimentation automatique
- Pas d'archivage

Exemple d'outil

Nom de la ressource	Python LMF library
Description	The Python LMF library is a suite of open-source Python modules for dictionary format conversion. It performs automatic tasks for multi-languages dictionaries, such as conversion between different formats used for dictionaries. The main idea of pylmflib is to provide a software package which integrates conversion functions from MDF format to several output formats: LaTeX (PDF), docx, HTML, etc. pylmflib implements the LMF standard. For more details, please see http://www.lexicalmarkupframework.org.
URL	https://pypi.python.org/pypi/pylmflib/1.0
Projet associé	ANR HimalCo
Licence	GPL
Droits d'accès	Téléchargement
Type d'outil	Bibliothèque logcielle
Environnement	Multi-plateforme
Langages de développement	Python
Interface graphique	Non
Formats d'entrée	MDF (toolbox), XML LMF
Formats de sortie	LaTeX (PDF), docx, HTML, etc.
Commentaires	Module Python présent sur le Python Package Index (PyPI). Installation avec pip. pylmflib a été écrit en Python 2.7.5.

Exemple de corpus

Nom de la ressource	Phono-LeoCola_Yamaguchi
Description	Etude longitudinale des productions spontanées d'un enfant de 15 mois à 5 ans
Projet associé	ANR Léonard / ANR Colaje
Droits d'accès	Téléchargement
Objectifs scientifiques	Base de données d'acquisition + étude de l'acquisition de la communication langagière
Modalité	Oral
Type de données	Corpus
Provenance des données	Productions spontannées d'un enfant (15 mois à 5 ans)
Formats de fichiers	MOV
Taille des données (Mo)	37 heures
Langue(s)	French
Types d'informations linguistiques	Orthographique, phonétique cible, phonétique produite, gestes (parfois)
État d'avancement	Achevé
Commentaires	Utilisation des logiciels Phon et Excel.

Perspectives : Consortium CORLI

- Réseau social
- Adaptable aux besoins
- Import de descriptions depuis des bases de données existantes : OLAC, Corpora, etc.
- Interface « user-friendly » : comptes personnels, badges, etc.
- Fédérer une communauté autour de la plateforme
- Gestion de commentaires
- Interactivité : façon Stack Overflow