

Produire des metadonnées au format Clarín

B. Gaiffe

2 décembre 2011

Le processus

Produire des
metadonnées
au format
Clarin

B. Gaiffe

- concevoir un profil de metadonnées
- obtenir un schéma XML pour ce profil
- deux cas :
 - étiqueter des données (cas 1)
 - traduire des metadonnées existantes (cas 2)

Concevoir ou réutiliser un profil de metadonnées

Les composants et profils sont spécifiés dans un langage particulier : CMDI.

Dans ce langage, on spécifie en termes logiques :

- la structure des composants (nom des composants, nom des éléments, types des éléments, ...)
- les cardinalités (nombres d'occurrences de sous-composants et d'éléments)
- les liens avec ISOCAT (ancrage des éléments)

Les moyens de le faire

- directement sur le site WEB : une interface permet de glisser/déposer des sous-composants, etc.
- en éditant un document XML valide en regard d'un schéma particulier (CMDI.xsd)

Attention : dans la seconde solution, il faudra enregistrer ses composants

Transformer la spécification en schéma XML

Pour éditer des metadonnées, il faut un schéma.
Le document CMDI de l'étape précédente n'est qu'une spécification.

Pour obtenir un schéma :

- passer par le site WEB
- ou, utiliser un ensemble de feuilles de style

Pour ceux qui connaissent, c'est la même démarche qu'en TEI...

Editer des metadonnées

A partir du schéma obtenu à l'étape précédente, on peut :

- editer des métadonnées pour des documents dans un éditeur XML généraliste (emacs, Oxygen, XML editor...)
- utiliser un éditeur “dédié” Arbil (remarquable pour l'édition “en masse”)
- transformer des métadonnées pré-existantes pour en obtenir une vue (partielle) en CMDI

Nos expériences en la matière

Produire des
metadonnées
au format
Clarin

B. Gaiffe

- ① corpus en TEI avec entpète TEI minimale (Est Républicain)
- ② corpus oraux (projet TCOF)

- ① encodé en TEI (entête assez minimal)
- ② format transcriber accompagné à l'origine de metadonnées en word.

- ① définition de composants et d'un profil *ad hoc*
- ② notre seule excuse : c'était il y a déjà longtemps
- ③ feuilles xsl pour projeter l'entête TEI sur métadonnées Clarin

Bilan :

- à refaire : complètement idiosyncrasique
- avec vous tous ? c'est-à-dire :
 - quelques composants partageables ?
 - génériques...
 - que nous repersonaliserons pour l'édition (nos métadonnées sont pauvres)

- 1 Définition avec les chercheurs(euses) d'un ensemble d'éléments/composants communs
- 2 en s'inspirant fortement de OLAC, DC, IMDI
- 3 réalisation d'une application WEB pour enregistrer les metadonnées (dans une base de données)
- 4 export de la base dans un jargon XML
- 5 traduction en schema CMDI (profil et composants)
- 6 traduction des metadonnées en CMDI

Bilan :

- à refaire (probablement)
- mais probablement essentiellement du renommage
- et du réancrage dans ISOCAT