

## Réunion INTER-EXPLO du jeudi 22/09/2016

Présents : Loïc Liégeois, Christophe Parisse, Christian Chanard, Frédéric Landragin, Mai Hodac (visio), Sarra El Ayari, Béatrice Turpin, Carole Etienne

Objet Les métadonnées : reprise des formulaires déposés sur le wiki, comment articuler écrit et oral, utilisation d'un éditeur TEI comme oXygen , formation et prochaines dates

### 1) Les métadonnées des corpus écrits

Actuellement, pas de solution commune dans les unités en dehors des métadonnées d'archivage , dépend des disciplines, difficile à faire → demander le budget pour un stagiaire de 6 mois (~3000 euros) de Master 2 , si possible au deuxième semestre si la mise en place du budget le permet, s'orienter de préférence vers un documentaliste qui pourra apporter son aide pour harmoniser et unifier l'existant : connaissance indispensable de la TEI (sémantique)

Pour les catégories de l'écrit, il existe une liste d'items plus ou moins consensuels ==> genre textuel

Pour associer oral et écrit, transformer enregistrement en collecte.

### 2) Rappel sur les métadonnées

La solution retenue s'organise en modules avec **un premier niveau commun de métadonnées**, c'est ce niveau qui est proposé dans le formulaire disponible sur le wiki, ensuite des modules complémentaires viendront s'ajouter par domaine. Ils avaient été définis dans l'Ircom mais à repréciser dans interexplo : métadonnées corpus, métadonnées enregistrement/collecte, métadonnées langues, métadonnées sociolinguistiques, métadonnées annotations , métadonnées juridiques.

### 3) Organisation technique

Un fichier ODD détaillé reprend l'ensemble des éléments et propriétés, leur caractère obligatoire/optionnel, l'ordre, les contraintes d'utilisation, le vocabulaire contrôlé, ... et des exemples d'utilisation via les éléments <elementSpec>, <constraintSpec > et <constraint> ,<exemplum>

Avantage : Il constituera notre référentiel et centralisera toutes les informations, il est en TEI et diffusable

Inconvénient: Il ne peut pas être traité par Roma avec les balises <elementSpec> ou <constraint>

A partir de ce fichier, un traitement automatique permettra de lister chaque élément terminal qui constituera une rubrique du formulaire: elle pourra être unique/multiple, obligatoire/optionnelle, ... suivant les caractéristique de l'élément dans le fichier ODD : one/oneOrMore/ZeroOrMore , ...

Les propriétés des nœuds terminaux deviendront aussi des rubriques optionnelles ou obligatoires usage="req" avec une liste fermée de valeurs dans <valList> et <valItem> du fichier ODD .

Les éléments non terminaux et les propriétés à valeur fixe (value="") ne constituent pas des rubriques du formulaire mais seront des balises TEI du fichier final.

oXygen : les utilisateurs qui souhaitent utiliser oXygen pour saisir/modifier les métadonnées en mode auteur se chargeront de générer un schéma à partir du fichier Odd et le css approprié, le référentiel commun est le fichier ODD.

#### 4) Le formulaire

Le formulaire est en ligne sur le web avec un téléchargement en fin de saisie du fichier TEI qui sera stocké en local par l'utilisateur, il pourra être rechargé dans la même interface pour être modifié/dupliqué.

Demander une adresse de messagerie avant de se connecter et envoyer le document à l'adresse ?

A priori, une saisie en 4 étapes indépendantes devrait suffire avec un fil d'ariane et la possibilité de stocker le résultat dans l'état pour le reprendre ultérieurement sans avoir à recommencer.

Dans les listes fermées, un choix "autre" évitera de choisir n'importe quelle valeur, un champ libre sera disponible pour complément (restrictions TEI pour certaines rubriques).

Un bouton d'aide affichera l'exemple de la balise <exemplum> du fichier ODD.

Un pop-up sera disponible pour poser une question à qq interlocuteurs de Corli.

Sur chaque rubrique, un bouton TEI permettra d'afficher la correspondance en TEI de la rubrique, et Dublin Core pour les rubriques qui correspondent à des métadonnées Dublin Core.

Un export en Dublin Core sera disponible.

Par la suite, il y aura aussi un export en CMDI/IMDI.

Prévoir quelques exemples vidéo en camtasia ou autre par type de données ?

#### 5) Les rubriques du formulaire déposé sur le wiki (niveau 0)

Il reste à définir ce qui est obligatoire ...

Métadonnées collection : labo, contributeurs, révisions peu variable écrit/oral <revisionDesc>

Métadonnées document: transcription, ...

Métadonnées données brutes: enregistrement, écrit numérisé non annoté

Métadonnées corpus : unité dans le regroupement

Métadonnées projet : rhapsodie, orféo, comere, ...

Terme de "Responsable" vague mais si on ajoute légal fait peur → responsable scientifique=la personne à qui s'adresser si besoin

Mettre le handle du document le plus haut possible → attention pas toujours de handle !

Ajouter dans <revisionDesc> la date du fichier ODD et la date de saisie du formulaire en auto.

Ajouter au corpus deux articles biblio, l'un en français l'autre en anglais → non, module additionnel Corpus

Problème du champ availability/licence si plusieurs enregistrements et une transcription avec des conditions d'utilisation différentes : sur quel objet porte availability ??? rappeler dans meta les licences de l'enregistrement ???

Anonymisation: liste fermée pour le moment → guide d'anonymisation à prévoir ? parfois dans le guide d'annotations l'ajouter car indispensable pour la réutilisation par des outils (prévu initialement dans le module additionnel enregistrement ou annotation)

Liste fermée des rôles : ajouter ajouter data\_inputer et compiler : synthèse et déplacer rôle dans la partie enregistrement/transcription

Ajouter un guide de collecte, protocole de recueil et structuration → module additionnel contexte ou signal/enregistrement

Ajouter date de parution pour l'écrit

Remonter lieu + date de collecte + date de parution

Speakers ==> à l'écrit pas de correspondance, les mêmes informations sur l'auteur ne sont pas utiles, les dates suffisent à déterminer par exemple l'âge de l'auteur

TAILLE = nb de mots du texte ( vient du TAL ou analyse textuelle: quid des mots composés, ... ) et nb d'annotations (libre)

Une info manuelle/auto par couche d'annotation ==> A AJOUTER même à l'oral

#### 6) Prochaines dates

Atelier pratique aux journées Floral en mars, l'écrit ne sera pas encore traité (dates du stage)

Colloque JLC 2017 en juillet à Grenoble

Formation Archivage + Métadonnées, si possible écrit et oral, deuxième trimestre ?