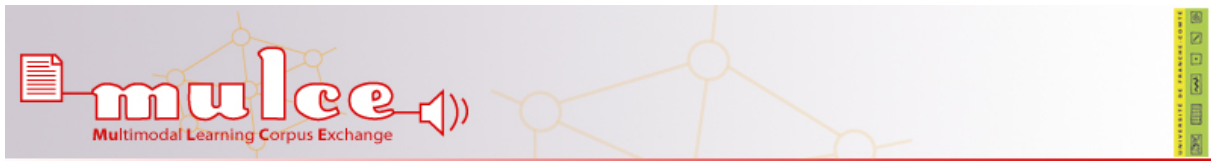


**ANR-06-CORP-006**

**Échange de corpus d'apprentissage multimodaux (MULCE)**



## **Rapport d'activité**

### **Tâche Tstruct1.5**

#### **Métadonnées de Mulce-struct**

Version du 28 novembre 2007 de TC, remise en page avec ajout synoptique en 2/08, ajout éléments location en mai 2013, TC

Mai 2013: Ce document est à utiliser avec circonspection car il ne fait qu'évoquer les discussions internes au projet. Il ne décrit pas les choix finaux, ni l'architecture complète des métadonnées Mulce.

#### **Auteur**

Marie-Laure Betbeder, LIFC

Thierry Chanier, LASELDI

## 1. Synopsis sous-tâche

Auteurs
Marie-Laure Betbeder , Thierry Chanier
Objectifs
<ul style="list-style-type: none"> <li>• Métadonnées pour corpus d'apprentissage et corpus distinguable, intégrées à Mulce-struct</li> <li>• Compatibilité avec normes et standards internationaux</li> <li>• Guidage sur la façon de déposer des corpus et de les renseigner</li> </ul>
Liens
Tpf1.1, Tdr1, Tstruct
Planning et état d'avancement
<p>Début : 1/11/07                  Fin : 31/03/08</p> <p>Il reste à déterminer la structure de métadonnées d'un corpus distinguable et à définir les champs obligatoires à renseigner lors du dépôt.</p>
Produit
<ul style="list-style-type: none"> <li>- Choix des composants permettant de définir toutes les informations sur un corpus d'apprentissage et à le rendre compatible avec les standards, à savoir:                         <ul style="list-style-type: none"> <li>o Dublin Core pour partie générale</li> <li>o Olac (<b>Open Language Archives Community</b> ) de façon à intégrer nos corpus d'apprentissage dans ce réseau international de corpus et à rendre le serveur Mulce comme serveur Olac, suivant le protocole OAI-OLAC pour la partage des données. Un nouveau type de corpus est proposé : learning &amp; teaching corpus</li> <li>o Lom (Learning Object Metadata) : pour la partie scénario pédagogique</li> </ul> </li> <li>- Construction de la structure propre à un corpus d'apprentissage, choix du préfixe "mce" pour nom d'espace XML et d'un répertoire pour dépôt des métas sur serveur Mulce.</li> </ul>
Annexes
Schémas XSD décrivant les métadonnées et fichiers XML de métadonnées pour Simuligne

## 2. Table des matières

1. Synopsis sous-tâche.....	2
2. Table des matières .....	3
3. Métadonnées des corpus et corpus distinguables : positionnement .....	5
3.1. Quoi référencer ?.....	5
3.2. Pourquoi référencer ?.....	5
3.3. Sur quels standards s'appuyer ?.....	5
3.4. Le format des métadonnées et leurs localisations .....	6
3.4.1. L'espace de nom et l'URL pour nos méta .....	6
3.5. La question des URN et de la nomenclature.....	6
3.5.1. Proposition CR et TC de début novembre .....	6
3.5.2. Nom anglais d'un corpus d'apprentissage .....	7
4. La LOM.....	8
5. Sélections de descripteurs Olac .....	9
5.1. Discourse Type .....	9
5.2. Linguistic subjects.....	9
5.3. Linguistic Data Type Vocabulary .....	10
5.4. Role vocabulary .....	10
5.5. Descripteur de lieux (coverage) et langue.....	11
5.5.1. Providing spatial coverage of a resource .....	11
5.5.2. Langues comme sujet d'étude ou d'usage dans le corpus.....	12
5.6. Extension OLAC.....	12
6. Dublin Core .....	14
6.1. Les schémas .....	14
6.2. Spécifier Noms personnes ou institutions .....	14
6.3. Les éléments qualifiés.....	14
6.4. Les formats.....	16
7. Les méta Mulce mce .....	17
8. L'empaquetage général et principe d'écriture des schémas .....	20
8.1. Vue d'ensemble sur fiche métadonnée d'un corpus d'apprentissage .....	20
8.2. Espaces de noms .....	20
8.3. Listage des fichiers et schémas pour un corpus d'apprentissage.....	20
8.3.1. Liste des schémas Mulce .....	20
8.3.2. Autres schémas .....	21
8.3.3. Fiches XML.....	21
8.4. Écriture des schémas .....	21
9. Ce qui reste à faire.....	22
10. Le serveur de dépôt Mulce en tant que serveur OAI - OLAC .....	23
11. Propositions.....	24
11.1. De Marie-Laure.....	24
11.1.1. Métadonnées des corpus.....	24

11.1.2. Corpus distinguable .....	26
11.2. Proposition venant de la tâche plate-forme .....	26
12. Annexes .....	28
12.1. Partie OLAC de la fiche de Simuligne.....	28
12.2. Partie Corpus de la fiche de Simuligne sans la LOM .....	30
12.3. Description LOM d'un scénario pédagogique.....	32
12.4. Entête fiche du corpus Simuligne avec espaces de noms.....	35
12.5. Annexe 1 : données Mulce venant d'autres sous-tâches.....	35
12.5.1. Traitement et accessibilité des données .....	35
12.5.2. Rôles .....	36
12.6. Annexe 2 : Descripteur corpus Clapi .....	36
12.7. Note CR après discussion CR / TC du 5/11 .....	38
12.7.1. Modification de Mulce-Struct : .....	38
12.7.2. Où et comment déclarer les contributeurs ?.....	39

### 3. Métadonnées des corpus et corpus distinguables : positionnement

#### 3.1. Quoi référencer ?

Il s'agit de créer des structures de métadonnées pour le corpus d'apprentissage et pour les corpus distinguables qui en sont extraits. Un corpus distinguable peut lui-même être décomposable en sous-corpus distinguables : l'ensemble des sessions d'un groupe Copeas et une session de ce groupe.

#### 3.2. Pourquoi référencer ?

En dehors des métadonnées qui seront intégrées au corpus ou parties de corpus, les fiches de chaque grand objet seront en ligne en tant que fichier XML. Elles indiqueront à tous les moissonneurs d'un coup ce que renferme nos corpus. Elles indiqueront également aux communautés ce que nous avons fait et ce qu'elles peuvent leur intéresser (OLAC, etc.).

Ces fiches devraient aider à avoir les points d'entrée pour le prototype Mulce dans son accès par corpus distinguable : voir doc Mulce Tpf1.1.

#### 3.3. Sur quels standards s'appuyer ?

Romary lors du colloque Epal a montré toute l'importance des standards en terme de métadonnées (en particulier). Ceux qui nous concernent ici sont :

- **LOM** : pour décrire des ressources pédagogiques. Même si un corpus d'apprentissage peut être utilisé comme une ressource pédagogique, ce n'est que par effet de bord. En revanche en examinant la liste des descripteurs, on peut trouver une bonne applicabilité au scénario pédagogique, autrement dit un scénario pédagogique peut devenir un ressource pédagogique (un objet distinguable au sens Mulce). D'autres formateurs peuvent avoir envie de l'utiliser (à noter qu'il n'existe pas encore de métadonnée pour les scénarios pédagogiques donc le choix Lom peut être justifié).
- **DC** : Le **Dublin Core** est plus qu'un standard, une norme destinée à identifier tout objet physique ou électronique. Nous devons l'utiliser au maximum. Les schémas sont en ligne.
- **OLAC** : **Open Language Archives Community**. Un programme qui organise de grosses bases de corpus, en accès libre, qui s'intéressent à toutes les langues du monde et peuvent donner lieu à tous types d'analyses. Un éditeur de corpus peut avoir son propre serveur. Il devra alors répondre au protocole OAI. Nous rassemblons des corpus d'apprentissage, notion qui n'existent pas pour eux, mais le contenu de nos *workspaces* et tous les documents rassemblés dans les formations sont des candidats à cette forme d'organisation. Le projet **Freebank** a construit une structure compatible avec Olac. A noter que les métadonnées de Olac sont définies comme des extensions du Dublin Core. Nombre de propriétés de métadonnées nous intéressent ici. Il faudra donc avoir une partie de notre structure en OLAC et réutiliser autant que faire se peut leur terminologie. Les schémas sont en ligne.
- **Clapi** : apparemment ce projet a fait sa structure sans se soucier de Olac, ce qui est assez aberrant. Il faudra examiner leur structure (voir annexe) pour voir si l'on oublie rien

Un corpus d'apprentissage a une structure particulière. Des éléments de cette structures doivent être documentés / référencés dans la structure de métadonnées. Une des premières question à se poser est : doit-on créer une structure particulière et la faire cohabiter avec les autres (Olac, etc.) ou doit-on faire des extensions aux standards pour y inclure nos informations.

La réponse peut se trouver pratiquement en identifiant les descripteurs dont on n'a besoin et de voir ensuite s'ils peuvent être intégrés aux standards, en ajoutant le cas échéant des extensions. Si beaucoup de descripteurs ne peuvent être aisément intégrés dans ces standards et si l'on pense que les descripteurs principaux d'un corpus d'apprentissage doivent être visible en tant que tels alors il faudra créer une structure propre.

### 3.4. Le format des métadonnées et leurs localisations

Elles seront en XML, dotés de schémas XSD, qui feront appel aux schémas des standards. Tous les noms des éléments, attributs, comme les valeurs fixées (pas les textes libres) devront être en anglais, uniquement.

Ces fichiers XML doivent pouvoir être isoler et manipuler en tant qu'objets individuels, de façon à pouvoir circuler hors sites Mulce. Mais il faudra sans doute aussi intégrer ces métadonnées à l'intérieur de la structure Mulce-struct (cf. nos graphiques anciens présentant Mulce-struct avec sa couche de métadonnées. Comment ? Cela est à discuter au plus tôt.

#### 3.4.1. L'espace de nom et l'URL pour nos méta

Il nous faut une URL pour nos espace de noms et pour mettre tous nos fichiers XML métadonnées.

Je serai pour : <http://mulce.univ-fcomte.fr/metadata> (voir <http://alsic.u-strasbg.fr/Metadata> )

### 3.5. La question des URN et de la nomenclature

Dans les métadonnées on doit utiliser des URI pour les descripteurs identifiants.. Nous allons utiliser dans ces métadonnées des URN et pas toujours des URL.

Nous constituons un corpus qui peut donc être vu comme un gros système documentaire avec des constituants qui ont besoin d'être identifiés de façon univoque :

- un corpus d'apprentissage X
- un corpus distinguable
- un scénario pédagogique d'un corpus X
- etc.

#### 3.5.1. Proposition CR et TC de début novembre

**Un corpus d'apprentissage (complet ou distinguable) dans mulce sera désigné par :**

**<institution>.<corpusname>.<couverture>.<constituant>.<format>.<version>**

exemple : mce.simu.all.ld.ims.v1

désignerait la version 1 (v1) du Learning Design (ld) au format IMS (ims) du corpus d'apprentissage (complet => couverture=all) de Simuligne (simu) issu de mulce (mce)

pour nos corpus, on aurait

- toujours "mce" pour <institution> ,
- "simu", "copeas", "tridem06", "tridem07" pour <corpusname>
- la couverture permet de distinguer le corpus d'apprentissage complet ("all") d'un corpus distinguable (par exemples : "forum", "aquiforum", "lugduposter" ou "chat" ou encore "Interculture")
- les <constituant> sont ceux définissant un corpus :
  - "all" pour tous
  - "ld" pour Learning Design
  - "rd" pour Research Design (ou Research Protocol pour employer l'ancienne formule)

"sid" pour Structured Interaction Data pour désigner l'instanciation "ric" pour Rights and Informed Consent pour la partie des licences "ana" pour Analysis
---

Figure 1 : nomenclature identifiant d'un corpus d'apprentissage

Le séparateur doit être le point (et non le double ::, voir les contraintes sur séparateurs de champs en Dublin Core, avec séparateur hiérarchiques : <http://dublincore.org/documents/dcmi-dcsv/> (DCMI DCSV: *A syntax for representing simple structured data in a text string*))

Ceci étant fixé, à partir de ce moment on s'en sert pour les métadonnées. Il va y avoir des identifiants (URI). Tous ne peuvent être des URL car on ne sait pas où sont localisés les objets précisément. Donc on met un URN. A partir de cet URN on peut formuler une requête qui va rechercher l'objet dans le BD Mulce. Donc ces URN figurent dans les champs de la BD (les identifiants de certains enregistrements). Ces URN vont servir aussi à tout le proto plate-forme pour les extractions, recherche.

De même "mce" sera le préfixe associé à tous les éléments et attributs XML de nos schémas pour les métadonnées, comme "dc" l'est pour le Dublin Core. Donc le préfixe de notre espace de nom *namespace*) Mulce.

Dans le même ordre d'idée nos URN seront fabriquées ainsi.

Les éléments devant pouvoir être repérés au niveau le plus haut dans les métadonnées par des URN sont :

- o 1 les membres (acteurs de la formation)
- o 2 les contributeurs (pour chaque partie)
- o 3 les plateformes (environnements de la formation)
- o - les documents de références pour les transcriptions (à revoir)

Propositions :

```
1 mce.simu.all.sid.memberlist
2 mce.simu.all.all.contributors => doit définir la liste des
contributeurs de la partie ???=ld/rd/sid/ric/ana
3 mce.simu.all.sid.platforms
```

### 3.5.2. Nom anglais d'un corpus d'apprentissage

Enfin, après discussion avec Marie-Noelle (3/11/07), nous proposons à Mulce d'adopter la terminologie suivante :

Au terme "corpus d'apprentissage" en français correspondra le terme "Learning and Teaching corpus" (L&T corpus)
---

Figure 2 : Nom en anglais d'un corpus d'apprentissage

*Learning corpus* ne passe pas car on risque de croire que c'est un corpus pour apprendre. *Learner corpus* non plus car cela ne concerne que les productions d'apprenants (pas autres participants).

*Learning and Teaching corpus*, dans cet ordre permet de mettre l'accent sur l'apprentissage en indiquant qu'il n'y a pas que des choses qui concernent les apprenants. Cela fait référence notamment à la présence de scénarios pédagogiques.

A valider en janvier avec tous les partenaires Mulce (voir aussi autres validation de terminologie) dans Tdr1.

Nous allons déclarer le type de corpus d'apprentissage comme un nouveau type linguistique dans OLAC (voir la section sur extensions OLAC).

## 4. La LOM

La LOM a été faite pour décrire des ressources pédagogiques, plutôt de taille réduite (un page Html, un exercice, un cours ou un peu plus -`aggregationlevel = 4`). Même si un corpus d'apprentissage peut être utilisé comme une ressource pédagogique, ce n'est que par effet de bord. En revanche la granularité et les descripteurs de la LOM permet de décrire un scénario pédagogique (et donc d'en faire une ressource pédagogique qui peut être considéré séparément du corpus (faire un fichier de méta à part et donner un accès dans l'interface, ce n'est pas un corpus distinguable mais un objet distinguable).

Le problème avec la LOM est le manque de vocabulaire de référence. Il suffit d'aller sur le site officiel pour voir le manque de sérieux à ce propos : à [http://www.imsglobal.org/metadata/mdv1p3/imsmd\\_bestv1p3.html](http://www.imsglobal.org/metadata/mdv1p3/imsmd_bestv1p3.html) vous constaterez que les URL des liens sur le vocabulaire contrôlé vont la plupart du temps sur rien ou sur des sites qui n'ont rien à voir. Ceci est inacceptable en termes de documentation (site non contrôlé) et cela montre l'état de la LOM !! Cependant dans les spécifications de la Lom on trouve quelques éléments généraux auquel on peut faire référence. Les canadiens et québécois ont fait un travail remarquable en utilisant la LOM et en développant leur propre vocabulaire, voir par exemple la ressource "L'autoévaluation", son fichier LOM : [http://fr.eureka.ntic.org/display\\_lo.php?format=LOM&lom\\_id=7869](http://fr.eureka.ntic.org/display_lo.php?format=LOM&lom_id=7869) et les vocabulaire associé, dont : <http://videos.cyberscol.qc.ca/xmlROA/vdex/vdex.910000.xml>, <http://videos.cyberscol.qc.ca/xmlROA/vdex/vdex.300000.xml>, <http://videos.cyberscol.qc.ca/xmlROA/vdex/vdex.100000.xml>

Nous utiliserons pour notre part la ressource de l'union européenne pour la taxinomie qui a le mérite de faire des correspondances entre plusieurs langues : European Treasury Browser (2005) le thésaurus : [http://www.eun.org/etb/thesaurus/alphadisplay\\_fr.rtf](http://www.eun.org/etb/thesaurus/alphadisplay_fr.rtf), accueil : <http://etb.eun.org/eun.org2/eun/fr/etb/content.cfm?lang=en&ov=7208>

Pour le fichier LOM décrivant le scénario Simuligne (`mce_simu_all_ld_lom_v1.xml`), voir en annexe. On notera l'utilisation des éléments:

- o `general`,
- o `lyfecycle` (utilisation `vcard` pour nom auteurs),
- o `technical` (référence à la partie plateforme Mulce : `mce.simu.all.sid.platforms`),
- o `educational` avec la `taxonpath` prenant ses valeurs dans ETB. J'ai choisi un ensemble de valeurs pour les éléments de la LOM retenu qui font partie du vocabulaire LOMv1.0, par exemple pour le public concerné par la formation / le scénario : `<context> = higher education`. On trouvera ce vocabulaire dans [http://www.imsglobal.org/metadata/imsmdv1p2p1/imsmd\\_bindv1p2p1.html](http://www.imsglobal.org/metadata/imsmdv1p2p1/imsmd_bindv1p2p1.html) : "IMS Learning Resource Meta-Data XML Binding "



## 5. Sélections de descripteurs Olac

Dans ce qui suit parlons de

- corpus d'apprentissage de type A s'il ne porte sur l'apprentissage des langues
- et de type B s'il porte sur l'apprentissage des langues.

Pour l'heure tous nos corpus sont de type B.

<http://www.language-archives.org/OLAC/1.0/olac.xsd>

### 5.1. Discourse Type

<http://www.language-archives.org/REC/discourse.html>

olac:discourse-type applies to type, subject (DC)

- \* drama
- \* formulaic\_discourse
- \* dialogue : Linguistic communication between two or more participants
- \* language\_play
- \* oratory
- \* narrative : A monologic discourse which represents temporally organized events
- \* procedural\_discourse
- \* report : A factual account of some event or circumstance.
- \* singing
- \* unintelligible\_speech

Essentiellement du type **dialogue** mais il peut aussi y avoir **report** (les rapports réflexifs) et **narrative** (pour les productions individuels des apprenants ?).

On met les 3 si l'objet référencé les contient.

### 5.2. Linguistic subjects

<http://www.language-archives.org/REC/field.html>

olac:linguistic-field applies to subject (DC)

<ul style="list-style-type: none"><li>* anthropological_linguistics</li><li>* applied_linguistics (The use of linguistic methods as applied to language and education, for example to literacy or language learning.)</li><li>* cognitive_science</li><li>* computational_linguistics</li><li>* discourse_analysis (The study of the patterns and meanings behind connected speech.)</li><li>* forensic_linguistics</li><li>* general_linguistics</li><li>* historical_linguistics</li><li>* history_of_linguistics</li><li>* language_acquisition</li></ul>	<ul style="list-style-type: none"><li>* morphology</li><li>* neurolinguistics</li><li>* philosophy_of_language</li><li>* phonetics</li><li>* phonology</li><li>* pragmatics</li><li>* psycholinguistics</li><li>* semantics</li><li>* sociolinguistics</li><li>* syntax</li><li>* text_and_corpus_linguistics (The study of the linguistic properties of an extended passage, text, or corpus of texts.)</li></ul>
--	--

<p>(Language Acquisition may be used to describe materials relating to either adult or child language acquisition, and to either first or later language acquisition. However, if the materials deal specifically with language teaching, or with the process of language learning from a pedagogical point of view, they may be best classified as Applied Linguistics.)</p> <ul style="list-style-type: none"> <li>* language_documentation</li> <li>* lexicography</li> <li>* linguistics_and_literature</li> <li>* linguistic_theories</li> <li>* mathematical_linguistics</li> </ul>	<ul style="list-style-type: none"> <li>* translating_and_interpreting</li> <li>* typology</li> <li>* writing_systems</li> </ul> <p>Note that a dataset should be described using the DCMI Type vocabulary, e.g. <b>Dataset, Collection.</b></p>
---	---

- o Pour les corpus B : `applied_linguistics`
- o Pour tous les corpus A et B : `discourse_analysis`
- o Pour tous les corpus A et B : `text_and_corpus_linguistics`

### 5.3. Linguistic Data Type Vocabulary

<http://www.language-archives.org/REC/type.html>

olac:linguistic-type applies to type (DC)

- \* lexicon
- \* primary\_text (Linguistic material which is itself the object of study, typically material in the subject language which is a performance of a speech event, or the written analog of such an event.)
- \* language\_description

- o Pour tous les corpus A et B : `primary_text` (même si cela contient de la vidéo)

C'est ici que nous étendrons le type Olac pour déclarer notre corpus d'apprentissage.

### 5.4. Role vocabulary

<http://www.language-archives.org/REC/role.html>

olac:role applies to creator et contributor (DC)

<ul style="list-style-type: none"> <li>* annotator</li> <li>* author</li> <li>* compiler (responsible for collecting the sub-parts of the resource together.)</li> <li>* consultant</li> <li>* data_inputter (responsible for entering, re-</li> </ul>	<ul style="list-style-type: none"> <li>* researcher (The resource was created as part of the participant's research, or the research presents interim or final results from the participant's research.)</li> <li>* research_participant (The participant acted as a research subject or responded to a questionnaire, the results of</li> </ul>
--	--

<p>typing, and/or structuring the data contained in the resource)</p> <ul style="list-style-type: none"> <li>* depositor (responsible for depositing the resource in an archive)</li> <li>* developer</li> <li>* editor (reviewed, corrected, and/or tested the resource)</li> <li>* illustrator</li> <li>* interpreter</li> <li>* interviewer</li> <li>* participant (was present during the creation of the resource, but did not contribute substantially to its content)</li> <li>* performer</li> <li>* photographer</li> <li>* recorder</li> </ul>	<p>which study form the basis of the resource.)</p> <ul style="list-style-type: none"> <li>* responder (The participant was an interlocutor in some sort of discourse event.)</li> <li>* signer</li> <li>* singer</li> <li>* speaker (The participant was a principal speaker in a resource that consists of a recording, a film, or a transcription of a recorded resource.)</li> <li>* sponsor (The participant contributed financial support to the creation of the resource)</li> <li>* transcriber (The participant produced a transcription of this or a related resource)</li> <li>* translator</li> </ul>
--	---

Ceci vient spécialisé sur terme DC contributor

```
< contributor xsi:type="olac:role">compiler</subject>
```

- o compiler et depositor -> collecteur corpus
- o author : ne pas utiliser sauf si auteur dudit document ou ressources
- o developer : CR pour prog ano, ingénieur pour proto plate-forme, Muriel pour V1
- o editor : ne pas utiliser sauf si travail correction, surtout ne pas confondre avec publisher.
- o interviewer : pour Copeas par exemple
- o participant : a priori **ne pas utiliser** et remplacer le terme utiliser dans Tdr1 par "acteur"
- o recorder : éventuellement pour MLB et CR durant séances Copéas, mais resercher peut suffire
- o researcher : les membres du projet Copeas, du projet Simuligne, Tridem donc y compris ceux qui ne font pas partie de Mulce.
- o responder : éventuellement, CR dans un entretien
- o speaker : les acteurs ; voir si utiliser aussi termino "actor"
- o sponsor : ANR, British Academy, ACI Odil.
- o transcriber : bien sûr pour transcripteurs

## 5.5. Descripteur de lieux (coverage) et langue

Ajout de mai 2013, d'après extrait document Ines Saddour

### 5.5.1. Providing spatial coverage of a resource

We indicate the spatial coverage of a resource in each metadata record. The elements used for the spatial characteristics are ISO3166 code for countries and TGN identifiers as far as cities are concerned.

For example, the spatial characteristics of a resource which is localized in Paris and Clermont-Ferrand in France would be indicated as follows:

```
<dcterms:spatial xsi:type="dcterms:ISO3166">FRANCE</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:TGN">7008202</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:ISO3166">FRANCE</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:TGN">7008038</dcterms:spatial>
```

### 5.5.2. Langues comme sujet d'étude ou d'usage dans le corpus

Sur les recommandation sde OLAC nous avons abandonné le format ISO639-2 pour prendre celui de OLAC (qui est bien dans ce standard aussi). Il faut égelement distinguer la langue dans l'élément language et dans l'élément subject

```
<dc:language xsi:type="dcterms:ISO639-2">eng</dc:language> == (changed
to)==> <dc:language xsi:type="olac:language" olac:code="eng"/>
And language (as a subject for teaching) have been added:
<dc:subject xsi:type="olac:language" olac:code="fra"/>
<dc:subject xsi:type="olac:language" olac:code="eng"/> -
Attention aux espaces (pourquoi ?) "URL:http://mulce.org"->"URL:
http://mulce.org"
<dc:subject xsi:type="dcterms:LCSH">multiple keywords</dc:subject> =>
multiple elements <dc:subject xsi:type="dcterms:LCSH">single
keyword</dc:subject>
```

## 5.6. Extension OLAC

La façon de faire des extensions est expliqué dans :

Ne pas oublier de décrire dans ce fichier ce qu'est un corpus d'apprentissage et le vocabulaire associé <http://www.language-archives.org/OLAC/metadata.html#Defining%20a%20third-party%20extension>

Cette extension doit être documentée et avoir une structure compatible avec <http://www.language-archives.org/OLAC/1.0/olac-extension.xsd>

Dans la figure 3 on décrit le schéma d'extension

```
<?xml version="1.0" encoding="utf-8"?>
<!-- Created with Liquid XML Studio 1.0.6.0 (http://www.liquid-
technologies.com) -->
<xs:schema xmlns:olac="http://www.language-archives.org/OLAC/1.0/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
attributeFormDefault="qualified" elementFormDefault="qualified"
xmlns:xs="http://www.w3.org/2001/XMLSchema">
  <xs:import
schemaLocation="http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd"
namespace="http://purl.org/dc/elements/1.1/" />
  <xs:import schemaLocation="olac.xsd" namespace="http://www.language-
archives.org/OLAC/1.0/" />
  <xs:annotation>
    <xs:appinfo>
      <olac-extension xmlns="http://www.language-
archives.org/OLAC/1.0/olac-extension.xsd">
        <shortName>linguistic-type2</shortName>
        <longName>Type for Learning and Teaching Corpus</longName>
        <versionDate>2007-11-15</versionDate>
        <description>Extension of the OLAC Linguistic Data Types for
Learning and Teaching Corpus</description>
        <appliesTo>type</appliesTo>
        <documentation>http://mulce.univ-
fcomte.fr/LTCorpus.html</documentation>
      </olac-extension>
    </xs:appinfo>
  </xs:annotation>
</xs:schema>
```

```

</xs:appinfo>
</xs:annotation>
<xs:complexType name="linguistic-type2">
  <xs:complexContent mixed="true">
    <xs:extension base="dc:SimpleLiteral">
      <xs:attribute name="code" type="mce-linguistic-type"
use="required" />
    </xs:extension>
  </xs:complexContent>
</xs:complexType>
<xs:simpleType name="mce-linguistic-type">
  <xs:union memberTypes="olac:olac-linguistic-type">
    <xs:simpleType>
      <xs:restriction base="xs:string">
        <xs:enumeration value="learning and teaching corpus" />
      </xs:restriction>
    </xs:simpleType>
  </xs:union>
</xs:simpleType>
</xs:schema>

```

Figure 3 : Extension Olac pour corpus d'apprentissage, fichier `mce_olac_ltcampus.xsd`

On remarque que l'on dit à quel élément du Dublin Core s'applique l'extension : ici à `type`. Cela étend `olac-linguistic-type` avec la valeur `learning and teaching corpus`.

Attention, il faudra rédiger et mettre en ligne un fichier d'explication sur ce qu'est un corpus d'apprentissage : <http://mulce.univ-fcomte.fr/LTCampus.html>

On l'utilise ainsi dans le fichier Xml méta dans la partie `olac` :

```

<dc:type xsi:type="linguistic-type2" code="learning and teaching corpus"
/>

```

Mon travail n'est pas tout à fait fini au niveau déclaration de nom d'espace. Il faudra que cela s'écrive ainsi :

```

<dc:type xsi:type="mce:linguistic-type" code="learning and teaching
corpus" />

```

Donc fichier schéma à modifier avec bons espace de noms.

## 6. Dublin Core

### 6.1. Les schémas

<http://dublincore.org/schemas/xmls/>

Target namespace Schema location

<http://purl.org/dc/elements/1.1/>

<http://dublincore.org/schemas/xmls/qdc/2006/01/06/dc.xsd>

<http://purl.org/dc/terms/>

<http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcterms.xsd>

<http://purl.org/dc/dcmitype/> <http://dublincore.org/schemas/xmls/qdc/2006/01/06/dcmitype.xsd>

### 6.2. Spécifier Noms personnes ou institutions

Voir l'élément detailed name d'un des schéma mulce : mce\_ltcorpus\_v1.xsd

```
<xs:complexType name="detailedName">
  <xs:complexContent mixed="true">
    <xs:extension base="dc:SimpleLiteral">
      <xs:sequence>
        <xs:element minOccurs="0" maxOccurs="unbounded" name="name" type="xs:string" />
      />
      <xs:element minOccurs="0" maxOccurs="1" name="institution" type="xs:string" />
      <xs:element minOccurs="0" maxOccurs="1" name="affiliationLocation" type="xs:string" />
      <xs:element minOccurs="0" maxOccurs="1" ref="mce:identifieur" />
    </xs:sequence>
  </xs:extension>
</xs:complexType>
```

### 6.3. Les éléments qualifiés

<http://dublincore.org/documents/dcmi-terms/>

DCMES Element	Element Refinement (s)	Element Encoding Scheme (s)	Utilisation Mulce		
<u>Title</u>	<u>Alternative</u>	-	oui		
<u>Creator</u>	-	-	Olac : collector et complier		
<u>Subject</u>	-	<table border="1"> <tr> <td><u>LCSH</u> <u>MeSH</u> <u>DDC</u> <u>LCC</u> <u>UDC</u></td> <td>Rameau Francis Pascal</td> </tr> </table>	<u>LCSH</u> <u>MeSH</u> <u>DDC</u> <u>LCC</u> <u>UDC</u>	Rameau Francis Pascal	Olac + Rameau et LCSH sur distance education, etc.
<u>LCSH</u> <u>MeSH</u> <u>DDC</u> <u>LCC</u> <u>UDC</u>	Rameau Francis Pascal				
<u>Description</u>	<u>Table Of Contents</u> <u>Abstract</u>	-	oui		
<u>Publisher</u>	-	-	Mulce		
<u>Contributor</u>	-	-			
<u>Date</u>	<u>Created</u> <u>Valid</u>	<u>DCMI Period</u> <u>W3C-DTF</u>	il s'agit du corpus et pas de la		

	<a href="#">Available Issued Modified Date Accepted Date Copyrighted Date Submitted</a>		formation created pour quand on l'a fait (le corpus) available pour quand on l'a mis en ligne YYY-MM-DD
<u>Type</u>	-	<a href="#">DCMI Type Vocabulary</a>	Collection et Dataset + Olac
<u>Format</u>	-	<a href="#">IMT</a> <a href="http://www.iana.org/assignments/media-types/">http://www.iana.org/assignments/media-types/</a>	oui, voir note ci-après
	<a href="#">Extent</a>	-	si video ou audio (durée) ou taille globale
	<a href="#">Medium</a>	-	
<u>Identifier</u>	-	<a href="#">URI</a>	nos URN ou URL
	<a href="#">Bibliographic Citation</a>	-	ppales références à nos publis
<u>Source</u>	-	<a href="#">URI</a>	
<u>Language</u>	-	<a href="#">ISO 639-2 RFC 3066</a>	fr ou en
<u>Relation</u>	<a href="#">Is Version Of</a> <a href="#">Has Version</a> <a href="#">Is Replaced By</a> <a href="#">Replaces</a> <a href="#">Is Required By</a> <a href="#">Requires</a> <a href="#">Is Part Of</a> <a href="#">Has Part</a> <a href="#">Is Referenced By</a> <a href="#">References</a> <a href="#">Is Format Of</a> <a href="#">Has Format</a> <a href="#">Conforms To</a>	<a href="#">URI</a>	has-part pour corpus distinguable avec sous partie conformsTo : pour IMSLD ?
<u>Coverage</u>	<a href="#">Spatial</a>	<a href="#">DCMI Point</a> <a href="#">ISO 3166</a> <a href="#">DCMI Box</a> <a href="#">TGN</a>	On est à distance mais on peut indiquer centre de rattachement ; Besançon, France ; Milton Keynes, GB
	<a href="#">Temporal</a>	<a href="#">DCMI Period</a> <a href="#">W3C-DTF</a>	name=Perth International Arts Festival, 2000; start=2000-01-26; end=2000-02-20 (pour DCMI)
<u>Rights</u>	<a href="#">Access Rights</a>	-	accessright pour encoder nos valeurs du param access

			Rights (ceux qui possède droits): Mulce avec Tdr1
	<a href="#">License</a>	<a href="#">URI</a>	URL pour nos licences d'utilisation
<a href="#">Audience</a>	<a href="#">Mediator Education Level</a>	-	Researcher or teachers in educational sciences ; linguistics ;
<a href="#">Provenance</a>	-	-	non
<a href="#">Rights Holder</a>	-	-	avec Tdr1
<a href="#">Instructional Method</a>	-	-	même problème que avec LOM
<a href="#">Accrual Method</a>	-	-	non
<a href="#">Accrual Periodicity</a>	-	-	non
<a href="#">Accrual Policy</a>	-	-	non

Les références à nos articles de recherche seront mises dans l'élément `BibliographicCitation`, de `identifier`. Pour tout corpus d'apprentissage ou distinguable, il faut absolument renseigner la partie références bibliographique car l'expérimentation a été fait avec un objectif de recherche: voir l'importance de cela dans CLAPI et CORINTHE et dans la base de corpus du LACITO.

Les références peuvent être mises en plein texte (format usuelle comme dans ce rapport), mais il serait bien de distinguer l'URL ou plutôt numéro OAI de nos références déposés dans des archives ouvertes (voir <http://dublincore.org/documents/dc-citation-guidelines/> ).

`info:oai/arXiv.org:hep-th/9901001`

Par exemple (extrait fiche Simuligne, partie Olac ) :

```
<dcterms:bibliographicCitation>Chanier, T. (2001).Créer des communautés d'apprentissage distance.
  Les dossiers de l'Ingénierie Educative, 36 (2001) 56-59 ;
  info:oai/edutice.archives-ouvertes.fr/edutice-00000119</dcterms:bibliographicCitation>
```

## 6.4. Les formats

Dans tous nos documents nous devons utiliser le standard IMT (<http://www.iana.org/assignments/media-types> ) pour les formats. Voir aussi

video/msvideo  
text/plain  
image/jpeg  
text/html  
application/msword  
cf

<http://www.commentcamarche.net/courrier-electronique/mime.php3>



## 7. Les méta Mulce mce

Typiquement ici il faut faire des structures pour

- corpus d'apprentissage
- Corpus distinguable
- Objet distinguable tel un scénario pédagogique
- Analyse / transcriptions

Dans la figure 4, on trouvera le début de description du corpus d'apprentissage Simuligne (ce n'est pas toute la fiche de métadonnées mais la partie en correspondance avec Mulce-struct, fichier `mce_simu_all_all_ltcampus_v1.xml` ). Le schéma correspondant est dans le fichier `mce_ltcampus_v1.xsd`. La partie LOM n'a pas été détaillée, elle figure sous forme d'entité générale `&lom;` et correspond au contenu précédemment cité.

**Cette façon de nommer et de structurer les choses est très importante. Il convient de la vérifier en détail, de la valider et d'étudier l'étroite correspondance avec Mulce-struct.** L'idée est que en récupérant les identifiants de cette structure on doit pouvoir accéder aux parties de la base de données Mulce-struct correspondante. A l'inverse on peut imaginer à terme qu'un tel type de fichier de métadonnées soit généré automatiquement par Mulce-struct.

Par exemple

- 1) Partie `learning-design`, ceci correspond au deux formats de description du scénario pédagogique :

```
<design id="mce.simu.all.ld.ims.v1" hasFormat="ims-ld" />
<design id="mce.simu.all.ld.motplus.v1" hasFormat="motplus"
/>
```

- 2) Dans la partie instanciation `<structured_interaction_data>`

On aurait accès la description des plates-formes, les éléments type et name donnant les types génériques et les noms de marque. Il faudrait écrire un type XSD pour ces types génériques.

```
<platforms>
  <identifiant>mce.simu.all.sid.platforms</identifiant>
  <type>vv</type>
  <name>vv</name>
</platforms>
```

- 3) On trouvera ici les dates de début et fin de la formation, alors que dans la partie LOM on met la durée.

```
<dc:type xsi:type="dcterms:DCMIType">Event</dc:type>
<dc:type>Course ; Research project</dc:type>
<dcterms:temporal xsi:type="detailedtemporal">
  <beginning_date>2001-04-09</beginning_date>
  <ending_date>2001-07-06</ending_date>
</dcterms:temporal>
```

Etc, etc. Voir en détail.

```
<corpus xmlns="http://mulce.univ-fcomte.fr/metadata"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xsi:schemaLocation="http://mulce.univ-fcomte.fr/metadata mce_ltcampus_v1.xsd" >
```

```

<description xml:lang="fr">Description du corpus d apprentissage Simuligne et de
ses sous-composants
</description>
<description xml:lang="en">xxx
</description>
<learning_design>
  <head_part_corpus>
    <identifiant>mce.simu.all.ld.v1</identifiant>
    <date>2007-07-01</date>
    <compiler>
      <contributor xsi:type="role2" code="compiler">
        <name>Noras, Muriel</name>
        <name>Chanier, Thierry</name>
        <institution>Universite de Franche-Comte</institution>
      </contributor>
    </compiler>
  </head_part_corpus>
  <course_design>
    <creator xsi:type="role2" code="course_designer">
      <name>Hassan, Xaviere</name>
      <name>Lamy, Marie-Noelle</name>
      <institution>The Open University</institution>
    </creator>
    <creator xsi:type="role2" code="course_designer">
      <name>Chanier, Thierry</name>
      <institution>Universite de Franche-Comte</institution>
    </creator>
    <contributor xsi:type="role2" code="developer">
      <name>Mbala, Aloys </name>
      <name>Dupont</name>
      <name>Kenab, Madani</name>
      <institution>Universite de Franche-Comte</institution>
    </contributor>
    <created>2001-01-01</created>
    &lom;
    <design id="mce.simu.all.ld.ims.v1" hasFormat="ims-ld" />
    <design id="mce.simu.all.ld.motplus.v1" hasFormat="motplus" />
  </course_design>
</learning_design>
<research_design>
  <head_part_corpus>
    <identifiant>mce.simu.all.rd.v1</identifiant>
    <description xml:lang="fr">xx
  </description>
    <description xml:lang="en">xx
  </description>
    <compiler>
      <contributor xsi:type="role2" code="compiler">zzz
    </contributor>
  </compiler>
    <date>2007-07-01</date>
  </head_part_corpus>
  <research_protocol>
    <creator xsi:type="role2" code="researcher">
      <name>Chanier, Thierry</name><name>Reffay, Christophe</name><name>Nicolet,
Jerome</name>
      <institution>Universite de Franche-Comte</institution>
    </creator>
    <creator xsi:type="role2" code="researcher">
      <name>Lamy, Marie-Noelle</name>
      <institution>The Open University</institution>
    </creator>
    <creator xsi:type="role2" code="researcher">
      <name>Trognon, Alain</name><name>Brixe Daniel</name>
      <institution>Universite Nancy 2</institution>
    </creator>
    <design id="mce.simu.all.rd.ims.v1" hasFormat="ims-ld" />
    <design id="mce.simu.all.rd.motplus.v1" hasFormat="motplus" />
  </research_protocol>
</research_design>
<structured_interaction_data>

```

```

<head_part_corpus>
  <identifiant>mce.simu.all.sid.v1</identifiant>
  <description xml:lang="fr">www
</description>
  <description xml:lang="en">www
</description>
  <compiler>
    <contributor>mce.simu.all.sid.contributors</contributor>
  </compiler>
</head_part_corpus>
<dc:type xsi:type="dcterms:DCMIType">Event</dc:type>
<dc:type>Course ; Research project</dc:type>
<dcterms:temporal xsi:type="detailedtemporal">
  <beginning_date>2001-04-09</beginning_date>
  <ending_date>2001-07-06</ending_date>
</dcterms:temporal>
<platforms>
  <identifiant>mce.simu.all.sid.platforms</identifiant>
  <type>vv</type>
  <name>vv</name>
</platforms>
<memberlist>
  <identifiant>mce.simu.all.sid.platforms</identifiant>
</memberlist>
<workspaces>
  <identifiant>mce.simu.all.sid.workspaces</identifiant>
</workspaces>
<observations>
  <identifiant>mce.simu.all.sid.obs</identifiant>ce qui a ete observee durant
formation</observations>
  <research_instantiation>
    <identifiant>mce.simu.all.sid.research</identifiant>
  </research_instantiation>
</structured_interaction_data>
<rights_informed_consent>
  <head_part_corpus>
    <identifiant>mce.simu.all.ric.v1</identifiant>
    <description xml:lang="fr">xxx
  </description>
    <description xml:lang="en">xxx
  </description>
  </head_part_corpus>
</rights_informed_consent>
<analysis>
  <head_part_corpus>
    <identifiant>mce.simu.all.ric.v1</identifiant>
  </head_part_corpus>
</analysis>
</corpus>

```

Figure 4 : Partie corpus d'apprentissage à la mulce de la fiche de métadonnées, soit le fichier mce\_simu\_all\_all\_ltcorpus\_v1.xml, voir aussi schéma mce\_ltcorpus\_v1.xsd

Je n'ai pas le temps de détailler plus aujourd'hui mais il y a beaucoup d'autres notions importantes : différences entre partie récolte de données `compiler` et la partie sur l'objet structurée, etc. Après discussion, avec vous je ferai une présentation au groupe de Besançon avec diapos.

## 8. L'empaquetage général et principe d'écriture des schémas

### 8.1. Vue d'ensemble sur fiche métadonnée d'un corpus d'apprentissage

Un fichier XML de métadonnée pour un corpus d'apprentissage sera composé des parties suivantes, voir figure 5 (et en annexe).

```

<mce:metadonnee > racine du document
<mce:header > métadonnée sur la fiche de métadonnée
<olac:olac> partie comprenant toute la structure Olac et Dublin Core avec
la déclaration de type corpus d'apprentissage, l'éditeur Mulce, les
publications.
  <dc:title xml:lang="fr">Corpus d'apprentissage Simuligne</dc:title>
  <dc:title xml:lang="en">Simuligne, Learning and Teaching
Corpus</dc:title>
...
</olac:olac>
<corpus> partie comprenant la structure d'un corpus propre à Mulce avec la
LOm pour e scénario pédagogique
  <learning_design>
  <research_design>
  <structured_interaction_data>
    <head_part_corpus>
    <platforms>
    <memberlist>
    <workspaces>
    <observations>
    <research_instantiation>
  <rights_informed_consent>
  <analysis>
</corpus>
</mce:metadonnee>

```

Figure 5 : extrait de la fiche du corpus Simuligne. `mce_simu_all_all_meta_v2.xml`

### 8.2. Espaces de noms

L'entête de la fiche du corpus Simuligne donne une idée de tous les espaces de nom associés et donc tous schémas auprès desquels valider les différentes parties :

```

<mce:metadonnee xmlns:mce="http://mulce.univ-fcomte.fr/metadata"
xmlns:olac="http://www.language-archives.org/OLAC/1.0/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:ims="http://www.imsglobal.org/xsd/imsmd_v1p2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://mulce.univ-fcomte.fr/metadata
mce_qualifieddc.xsd">

```

### 8.3. Listage des fichiers et schémas pour un corpus d'apprentissage

#### 8.3.1. Liste des schémas Mulce

- `mce_qualifieddc.xsd` : schéma de plus haut niveau qui appelle presque tous les autres
- `mce_ltcampus_v1.xsd` : schéma de la structure corpus
- `mce_olac_ltcampus.xsd` : schéma pour étendre Olac pour type corpus d'apprentissage.

### 8.3.2. Autres schémas

- dc.xsd : éléments simples Dublin Core
- dcmitype.xsd : types Dublin Core
- dcterms.xsd : éléments qualifiés Dublin Core (appelle les 2 précédents)
- imsmd\_v1p2p4.xsd : schéma LOM V1.2.4
- olac.xsd : schéma général Olac, appelle les autres
- olac-discourse-type.xsd
- olac-extension.xsd
- olac-language.xsd (pas utilisé)
- olac-linguistic-field.xsd
- olac-linguistic-type.xsd
- olac-role.xsd
- xhtml-lat1.ent : entités XHTML pour encoder accents, sinon risque de problèmes suivant plates-formes. Les accents n'ont pas encore utilisés mais le seront dans version finale
- xhtml-special.ent

### 8.3.3. Fiches XML

La fiche de métadonnées du corpus Simuligne est rassemblée dans un seul fichier :

```
mce_simu_all_all_meta_v2.xml
```

Mais elle a été constituée à partir des éléments suivants.

```
mce_simu_all_all_ltcorpus_v1.xml
mce_simu_all_all_olac_v1.xml
mce_simu_all_ld_lom_v1.xml
```

## 8.4. Écriture des schémas

Il peut être intéressant de développer chaque partie de fiche de corpus (3, 4 avec LOM) séparément, de les valider séparément, en particulier la LOM et ensuite de tout assembler et revalider avec les schémas.

Une remarque sur la façon de renseigner **TOUS nos schémas (ici et dans Mulce-struct)** La documentation d'un schéma doit d'abord se faire dans le schéma à l'aide le l'élément annotation (avant tut autre documentation extérieure). Toutes les structures importantes doivent être commentées dès la première version. Par exemple :

```
<xs:element name="corpus">
  <xs:complexType>
    <xs:annotation>
      <xs:documentation xml:lang="fr">Vue ensemble sur corpus
      apprentissage. La partie recherche est facultative. Tous les elements de
      nom "identifier" ou les attributs de nom "id" ont des contenus
      correspondant aux URN Mulce, c est a dire permettent d acceder a la partie
      correspondante dans la base Mulce</xs:documentation>
    </xs:annotation>
    <xs:sequence>
      <xs:element maxOccurs="unbounded" ref="mce:description" />
      <xs:element minOccurs="1" maxOccurs="1" ref="mce:learning_design" />
    </xs:sequence>
  </xs:complexType>
</xs:element>
```

Les annotations seront soit dans les deux langues (en et fr) , soit si manque de temps, **uniquement en anglais**. Ce n'est pas ce que j'ai fait jusqu'à présent mais c'est une erreur, donc à changer.

Tous les schémas Mulce ont été écrits avec le logiciel *Liquid* (2007) à la main. Mes compétences en schéma étant récentes, des maladresses, erreurs, oublis sont bien sûr possibles.

Certains types ont été développés ici pour les métadonnées mais peuvent nécessiter une harmonisation avec des parties de Mulce-struct (cf. roles dans Mulce, etc.).

## 9. Ce qui reste à faire

Malgré le temps passé (mais beaucoup d'inexpérience en schéma et de temps passé à découvrir les structures des autres) tout n'est pas fini, mais il y a aujourd'hui une structure complète de corpus d'apprentissage (avec quelques champs / éléments encore vides). La validation va être importante car :

- Grande proximité avec la structure, les URN de Mulce-struct et avec ce que sont demain les fiches produites pour les requêtes OAI posée auprès de notre serveur.
- Détermination d'une fraction notable de la terminologie que nous emploierons ensuite entre nous et dans nos articles et qui a été faite pour montrer notre ouverture aux standards.
- En 2008, il faudra d'ailleurs prévoir de nous déclarer auprès de l'Olac (voir aussi section sur serveur OAI).

Listons ce qui reste à faire :

- Validation par MLB et CR de la fiche de corpus Simuligne (toutes les sous-parties) et des schémas et types (cf. role, type fichiers, etc.)
- Remplissage complet de cette fiche avec langue française et anglaise
- Création fiche corpus distinguable
- Étude relation automatique entre la BD Mulce pour requête fiche vers Mulce et pour génération fiche à partir de Mulce
- Étude relation métadonnées actuelle avec celles pour notre serveur OAI-OLAC. Ce point rejoindra peut-être le précédent.

Je ne serai pas en mesure d'accomplir tout ce travail. Voir ce qui sera du ressort des ingénieurs et techniciens embauchés.

## **10. Le serveur de dépôt Mulce en tant que serveur OAI - OLAC**

Les corpus de type OLAC sont déposés dans des serveurs de dépôt nécessairement compatibles OAI.

La proposition est donc de déclarer le serveur Mulce comme un serveur OLAC. Dans ce document on trouve la description OAI et les serveurs dynamiques ou statiques (à creuser).

<http://www.language-archives.org/OLAC/repositories.html>

## 11. Propositions

### 11.1. De Marie-Laure

**Mai 2013 : Partie non retenue**

#### 11.1.1. Métadonnées des corpus

##### ➤ L'exemple de Clapi

Dans les corpus à feuilleter de Clapi, les échantillons de corpus sont présentés à l'aide de quelques métadonnées :

Chaque corpus a un nom et une présentation succincte avec les champs suivants :

##### Métadonnées du corpus :

Nom du champ	Type de données	« Exemple » ou précision
Nom	Texte	« ADI Etudiants »
Collecté	Texte	Date+ contexte (cours X, jour de la semaine, heure)
Présentation	Texte	« Afficher » = Lien, signification ?
Responsable	Texte	
Durée	Durée	« 01 :30 :30 »
Langue	Texte	
Nb locuteurs	Nombre	
NB enregistrements	Nombre	« 2 enregistrement(s) »

Suivent une présentation des enregistrements eux-mêmes associés à une ou plusieurs transcriptions.

##### Métadonnées d'un enregistrement :

Nom	Type de données	« Exemple » ou précision
Durée	Durée	
Recueil	Texte	« visible »
Genre détaillé	Texte	« Débat d'étudiants sur l'immigration »
Genre interactionnel	Texte	
Collecteur	Texte	1 ou plusieurs noms
Présentation	Texte	
Accès	Texte	« en libre accès »



Nb transcription(s)	nombre	
Lieu	Texte	Ville
Support	Texte	Audio, nom anonymisé, de bonne qualité

Les deux enregistrements proposés semblent être un extrait pour le premier et l'enregistrement intégral pour le second. Leur accessibilité varie (soumis à la signature d'une convention de recherche pour l'enregistrement intégral).

Il est possible d'afficher les locuteurs, on accède aux informations suivantes : code du locuteur, sexe, année de naissance, âge au moment de l'expérimentation, statut, études (par exemple supérieur) et langue (« maternelle : français »).

Possibilité de jouer un extrait en streaming.

### Métadonnées d'une transcription :

Nom	Type de données	« Exemple » ou précision
Type	Durée	« Totale en orthographe adaptés »
Convention	Texte	« Afficher » Lien vers le fichier
Commentaire	Texte	« transcription en format liste, passé en convention ICOR »
Accès	Texte	« En libre accès -> Télécharger la transcription
Traitement	Texte	« Toilettée minutée »
Format	Texte	« Informatique

Différence entre les 2 transcriptions : le champ commentaire : transcription en format liste, passé en convention ICOR et transcription en partition.

#### ➤ Proposition pour Mulce

##### Corpus principal

Nom	Type de données	« Exemple » ou précision
Nom du corpus	Texte	
Contexte de formation	Texte	
Responsable	Texte	
« Personnes participantes »	Texte	
Présentation succincte	Texte	
Durée	durée	
Langue principale	Texte	

Support	Texte	
Nombre de participants	Nombre	
Environnement technologique	Texte	

### 11.1.2. Corpus distinguable

Définition d'un corpus distinguable :

**A discuter :** Un corpus distinguable est une sélection d'un corpus. Cette sélection peut être **longitudinale ou transversale** : les interactions/productions d'un groupe, une des séances d'une formation ou un ensemble d'interactions particulier comme les actes de forum d'un groupe. Un corpus distinguable n'est **téléchargeable** (?) qu'avec le scénario pédagogique et l'instanciation.

**A discuter :** Exemple pour Simuligne : 4 corpus distinguable (4 groupes), chacun pouvant être décomposé en 3 corpus distinguable : les interactions d'un type d'outil de communication.

**A discuter :** Exemple pour Copéas : 2 corpus distinguable (2 groupes), chacun pouvant être décomposé en 8 corpus distinguable (1 par séance).

**A discuter :** Attention : il semble difficile de définir des corpus distinguables qui se recoupent (un corpus distinguable peut en englober un).

Nom	Type de données	« Exemple » ou précision
Nom du corpus principal	Texte	
Nom du corpus distinguable	Texte	
Contexte de formation	Texte	
Responsable	Texte	
« <b>Personnes participantes</b> »	Texte	
Durée	durée	
Langue principale	Texte	
Support	Texte	

### 11.2. Proposition venant de la tâche plate-forme

La fiche descriptive d'un corpus contient :

- nom du corpus
- date de collecte
- nom du responsable de corpus
- durée
- auteur
- contenu
- langue
- plate forme de formation

Ainsi qu'un bref descriptif de chaque enregistrement :

- durée
- genre
- collecteur
- présentation
- accès
- lieu
- support
- qualité
- extrait

Et aussi des transcriptions :

- type
- convention
- accès
- traitement
- format

**Corpus distinguable** : Une fiche de métadonnées remplie par le responsable du corpus décrit un corpus distinguable.

Méta données à renseigner :

- Nom de l'auteur
- Date
- Nom du fichier
- Description

A ces méta données seront ajoutées celles plus générales contenues dans les méta données du corpus

ANALYSE :

Toute analyse déposée devra être accompagné par une fiche descriptive contenant :

- des métadonnées
  - ce qui diffère par rapport à la version précédente (si on part d'une analyse existante)
  - le type d'analyse réalisée
  - les auteurs
  - la date de dépôt
  - la version (numéro)
  - le nom
- des liens vers le(s) fichier(s) de données de départ / analyse.

## 12. Annexes

### 12.1. Partie OLAC de la fiche de Simuligne

Mise à jour en mai 2013

```

<oai:record>
  <oai:header>
    <oai:identifiant xmlns="http://www.openarchives.org/OAI/2.0/oai-identifiant"
xsi:schemaLocation=" http://www.openarchives.org/OAI/2.0/oai-identifiant
http://www.openarchives.org/OAI/2.0/oai-identifiant.xsd"
>oai:mulce.org:mce.simu.all.all</oai:identifiant>
    <oai:datestamp>2012-07-17</oai:datestamp>
  </oai:header>
  <oai:metadata>
    <olac:olac xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:olac="http://www.language-archives.org/OLAC/1.1/"
xmlns:mce="http://lrl-diffusion.univ-bpclermont.fr/mulce/metadata"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation=" http://purl.org/dc/elements/1.1/ http://www.language-
archives.org/OLAC/1.1/dc.xsd http://purl.org/dc/terms/ http://www.language-
archives.org/OLAC/1.1/dcterms.xsd http://www.language-archives.org/OLAC/1.1/
http://www.language-archives.org/OLAC/1.1/olac.xsd http://lrl-
diffusion.univ-bpclermont.fr/mulce/metadata http://lrl-diffusion.univ-
bpclermont.fr/mulce/metadata/mce-sc
hemas/mce_letec_meta.xsd">
    <dc:title xml:lang="fr">Corpus d'apprentissage Simuligne</dc:title>
    <dc:title xml:lang="en">LETEC (Learning and Teaching Corpus)
Simuligne</dc:title>
    <dc:description xml:lang="fr">Corpus d'apprentissage de la formation en ligne
Simuligne (2001), dont le scénario est basée sur une simulation globale pour
l'apprentissage du français langue étrangère (FLE) et inclut également une
étape interculturelle &quot;Interculture&quot; inspirée de Cultura. Ce
corpus comprend le scénario pédagogique dans plusieurs formats, le protocole
de recherche (et ses données recueillies), les interactions en ligne et les
productions des apprenants structurées suivant un schéma XML, la liste des
participants, les licences d'utilisation.</dc:description>
    <dc:description xml:lang="en">This is the Learning and Teaching Corpus of the
online educational experiment Simuligne (2001). Its scenario is based on a
global simulation for the learning of French as a foreign language. It also
includes an intercultural activity, &quot;Interculture&quot;, based on the
Cultura project. The corpus includes the pedagogical scenario, described in
several formats, the research protocol, participant's online interactions
and productions (structured in XML), list of participants, licences of
use.</dc:description>
    <dc:description xml:lang="en">Analyses done by researchers on this LETEC
corpus can be found on Mulce Website (http://mulce.org) or Mulce repository
(http://repository.mulce.org) . They are contained in distinguishable
corpora.</dc:description>
    <dc:description xml:lang="en"> This corpus contains a total of 11638 acts, of
which, 6790 are chat acts, 2686 forum acts, 2030 email acts and 132
production acts.</dc:description>
    <dc:creator olac:code="compiler" xsi:type="olac:role">Chanier, Thierry ;
Lamy, Marie-Noelle ; Reffay, Christophe</dc:creator>
    <dc:contributor>Reffay, Christophe ; Chanier, Thierry ; Betbeder, Marie-
Laure</dc:contributor>
    <dc:contributor olac:code="sponsor" xsi:type="olac:role">ANR-06-CORP-006
echange de corpus d'apprentissage multimodaux (MULCE), Programme
&quot;CORPUS ET OUTILS DE LA RECHERCHE EN SCIENCES HUMAINES ET

```

```

SOCIALES" ; 2006, Agence Nationale de la Recherche,
France</dc:contributor>
<dc:contributor olac:code="sponsor" xsi:type="olac:role">Projet de recherche
Icogad, programme Cognitique 2000-2002, Ministre de la Recherche,
France.</dc:contributor>
<dc:publisher>Mulce (MULTimodal Corpus Exchange) ; Universite Blaise Pascal ;
Clermont-Ferrand:France ; http://mulce.org</dc:publisher>
<dcterms:created xsi:type="dcterms:W3CDTF">2009-04-20</dcterms:created>
<dc:language xsi:type="olac:language" olac:code="eng"/>
<dc:language xsi:type="olac:language" olac:code="fra"/>
<dc:subject xsi:type="olac:language" olac:code="fra"/>
<dc:subject xsi:type="dcterms:LCSH">Education</dc:subject>
<dc:subject xsi:type="dcterms:LCSH">Data processing</dc:subject>
<dc:subject
xsi:type="dcterms:LCSH">Computer-assisted
instruction</dc:subject>
<dc:subject xsi:type="dcterms:LCSH">Language and languages</dc:subject>
<dc:subject xsi:type="dcterms:LCSH">Study and teaching</dc:subject>
<dc:subject olac:code="applied_linguistics" xsi:type="olac:linguistic-
field"/>
<dc:subject olac:code="discourse_analysis" xsi:type="olac:linguistic-field"/>
<dc:subject olac:code="text_and_corpus_linguistics"
xsi:type="olac:linguistic-field"/>
<dc:subject xml:lang="fr">Le corpus d'apprentissage Simuligne est un ensemble
de données structurées contenant plusieurs composants décrivant
respectivement : le scénario pédagogique, le protocole et les questions de
recherche, les interactions des acteurs dans les
environnements.</dc:subject>
<dc:subject xml:lang="en">The learning and teaching corpus named Simuligne is
a set of structured data containing various interconnected components:
learning design, research questions and protocol, actors interaction data
coming from virtual environments and right informed consents.</dc:subject>
<dc:type xsi:type="dcterms:DCMIType">Dataset</dc:type>
<dc:type xsi:type="dcterms:DCMIType">Collection</dc:type>
<dc:type olac:code="primary_text" xsi:type="olac:linguistic-type"/>
<dc:type mce:code="learning and teaching corpus" xsi:type="mce:linguistic-
type2"/>
<dc:type olac:code="dialogue" xsi:type="olac:discourse-type"/>
<dc:type olac:code="narrative" xsi:type="olac:discourse-type"/>
<dc:format xsi:type="dcterms:IMT">text/html</dc:format>
<dc:format xsi:type="dcterms:IMT">text/xml</dc:format>
<dc:format xsi:type="dcterms:IMT">text/rtf</dc:format>
<dc:format xsi:type="dcterms:IMT">application/pdf</dc:format>
<dc:format xsi:type="dcterms:IMT">application/msword</dc:format>
<dc:format xsi:type="dcterms:IMT">image/jpeg</dc:format>
<dc:format xsi:type="dcterms:IMT">image/gif</dc:format>
<dc:format xsi:type="dcterms:IMT">audio/x-wav</dc:format>
<dcterms:extent>26000 ko</dcterms:extent>
<dcterms:extent>11638 acts ; 345875 tokens</dcterms:extent>
<dcterms:temporal>name=Simuligne course ; start=2001-04-09; end=2001-07-
06</dcterms:temporal>
<dcterms:spatial xsi:type="dcterms:ISO3166">GB</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:TGN">7026232</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:ISO3166">FR</dcterms:spatial>
<dcterms:spatial xsi:type="dcterms:TGN">7008356</dcterms:spatial>
<dcterms:conformsTo>IMS-LD for learning design and research protocol ; IMS-CP
for packaging ; Mulce-struct for the SID subpart</dcterms:conformsTo>
<dcterms:conformsTo xsi:type="dcterms:URI"
>http://lrl-diffusion.univ-
bpclermont.fr/mulce/metadata/mce_LETECorpus-en.pdf</dcterms:conformsTo>
<dcterms:references>For Interculture : Cultura projet,
http://web.mit.edu/french/culturaNEH/</dcterms:references>
<dcterms:references>For the global simulation : Yaiche, Francis (1996). Les
simulations globales, mode d'emploi. Hachette FLE :
Paris</dcterms:references>

```

```

<dcterms:hasPart
xsi:type="mce:idTypeMulce">mce.simu.all.sid</dcterms:hasPart>
<dcterms:hasPart xsi:type="mce:idTypeMulce">mce.simu.all.ld</dcterms:hasPart>
<dcterms:hasPart xsi:type="mce:idTypeMulce">mce.simu.all.rp</dcterms:hasPart>
<dcterms:hasPart
xsi:type="mce:idTypeMulce">mce.simu.all.ric</dcterms:hasPart>

<dc:identifiant xsi:type="dcterms:URI" >http://mulce.univ-
bpclermont.fr:8080/PlateFormeMulce/VIEW/PUBLIC/03/VMeta.do?adr=Simuligne%2FCo
rpus_objets%2Fmce.simu.all.all-CP</dc:identifiant>
<dc:identifiant xsi:type="mce:idTypeMulce">mce.simu.all.all</dc:identifiant>
<dcterms:bibliographicCitation>Reffay, C. Chanier, T. Lamy, M.-N. &
Betbeder, M.-L. (2009). (editors). LETEC corpus Simuligne. Mulce.org :
Clermont Université. [oai : mulce.org:mce.simu.all.all ;
http://repository.mulce.org ] </dcterms:bibliographicCitation>

<dcterms:references>Chanier, T. (2001).Créer des communautés d'apprentissage
à distance. Les dossiers de l'Ingénierie Educative, 36 (2001) 56-59 ;
info:oai/edutice.archives-ouvertes.fr/edutice-00000119</dcterms:references>
<dcterms:references>Reffay, C. , Chanier, T.: "How social network
analysis can help to measure cohesion in collaborative distance-
learning";, Proceeding of Computer Supported Collaborative Learning
conference (CSCL'2003), June, Bergen, Norway. Kluwer Acedemic Publisher. pp
343-352 ; info:oai/edutice.archives-ouvertes.fr/edutice-
00000422</dcterms:references>
<dcterms:references>Reffay, C, Chanier, T., Noras, M. and Betbeder, M.-L.
(2008). Contribution à la structuration de corpus d'apprentissage pour un
meilleur partage en recherche. In Basque, J. and Reffay, C. (dir.), numéro
spécial EPAL (échanger pour apprendre en ligne), Sciences et Technologies de
l'Information et de la Communication pour l'Education et la Formation
(STICEF), 15, [http://sticef.univ-lemans.fr/num/vol2008/01-
reffay/sticef_2008_reffay_01p.pdf , http://edutice.archives-
ouvertes.fr/edutice-00159733 ]</dcterms:references>
<dcterms:references> Lamy, M-N. (2006) 'Interactive Task Design and the Whole
Learner' In Garcia Mayo, P. (ed.) Investigating Tasks in Formal Language
Settings, Multilingual Matters. 242-264.</dcterms:references>
<dcterms:references> Lamy M-N. and Hassan X.P. (2003) 'What influences
reflective interaction in distance peer learning? Evidence from four long-
term online learners of French', in Open Learning Journal, Vol 18, n° 1, pp
39-59</dcterms:references>
<dcterms:references xsi:type="dcterms:URI" >http://edutice.archives-
ouvertes.fr/edutice-00000119</dcterms:references>
<dcterms:references xsi:type="dcterms:URI" >http://edutice.archives-
ouvertes.fr/edutice-00000422</dcterms:references>
<dcterms:references xsi:type="dcterms:URI" >http://edutice.archives-
ouvertes.fr/edutice-00159733</dcterms:references>
<dc:rights xsi:type="dcterms:URI" >http://lrl-diffusion.univ-
bpclermont.fr/mulce/metadata/vdex/mce_licence.xml</dc:rights>
<dcterms:accessRights>open access after registration</dcterms:accessRights>
<dc:rights>Rights holders of this corpus are: Thierry Chanier ; Marie-Noelle
Lamy ; Christophe Reffay ; Marie-Laure betbeder ; Maud Ciekanski</dc:rights>
<dc:rights>Creative Common License: http://creativecommons.org/licenses/by-
nc-sa/2.0/</dc:rights>
<dcterms:audience>Researcher or teachers in educational sciences or
linguistics</dcterms:audience>
</olac:olac>
</oai:metadata>
</oai:record>

```

## 12.2. Partie Corpus de la fiche de Simuligne sans la LOM

```
<corpus xmlns="http://mulce.univ-fcomte.fr/metadata">
```

```

<description xml:lang="fr">Description du corpus d apprentissage Simuligne et de
ses sous-composants
</description>
<description xml:lang="en">xxx
</description>
<learning_design>
  <head_part_corpus>
    <identifiant>mce.simu.all.ld.v1</identifiant>
    <date>2007-07-01</date>
    <compiler>
      <contributor xsi:type="role2" code="compiler">
        <name>Noras, Muriel</name>
        <name>Chanier, Thierry</name>
        <institution>Universite de Franche-Comte</institution>
      </contributor>
    </compiler>
  </head_part_corpus>
  <course_design>
    <creator xsi:type="role2" code="course_designer">
      <name>Hassan, Xaviere</name>
      <name>Lamy, Marie-Noelle</name>
      <institution>The Open University</institution>
    </creator>
    <creator xsi:type="role2" code="course_designer">
      <name>Chanier, Thierry</name>
      <institution>Universite de Franche-Comte</institution>
    </creator>
    <contributor xsi:type="role2" code="developer">
      <name>Mbala, Aloys </name>
      <name>Dupont</name>
      <name>Kenab, Madani</name>
      <institution>Universite de Franche-Comte</institution>
    </contributor>
    <created>2001-01-01</created>
    &lom;
    <design id="mce.simu.all.ld.ims.v1" hasFormat="ims-ld" />
    <design id="mce.simu.all.ld.motplus.v1" hasFormat="motplus" />
  </course_design>
</learning_design>
<research_design>
  <head_part_corpus>
    <identifiant>mce.simu.all.rd.v1</identifiant>
    <description xml:lang="fr">xx
  </description>
    <description xml:lang="en">xx
  </description>
    <compiler>
      <contributor xsi:type="role2" code="compiler">zzz
    </contributor>
    </compiler>
    <date>2007-07-01</date>
  </head_part_corpus>
<research_protocol>
  <creator xsi:type="role2" code="researcher">
    <name>Chanier, Thierry</name>
    <name>Reffay, Christophe</name>
    <name>Nicolet, Jerome</name>
    <institution>Universite de Franche-Comte</institution>
  </creator>
  <creator xsi:type="role2" code="researcher">
    <name>Lamy, Marie-Noelle</name>
    <institution>The Open University</institution>
  </creator>
  <creator xsi:type="role2" code="researcher">
    <name>Trognon, Alain</name>
    <name>Brixe Daniel</name>
    <institution>Universite Nancy 2</institution>
  </creator>
  <design id="mce.simu.all.rd.ims.v1" hasFormat="ims-ld" />
  <design id="mce.simu.all.rd.motplus.v1" hasFormat="motplus" />
</research_protocol>

```

```

</research_design>
<structured_interaction_data>
  <head_part_corpus>
    <identifieur>mce.simu.all.sid.v1</identifieur>
    <description xml:lang="fr">www
  </description>
    <description xml:lang="en">www
  </description>
    <compiler>
      <contributor>mce.simu.all.sid.contributors</contributor>
    </compiler>
  </head_part_corpus>
  <dc:type xsi:type="dcterms:DCMIType">Event</dc:type>
  <dc:type>Course ; Research project</dc:type>
  <dcterms:temporal xsi:type="detailedtemporal">
    <beginning_date>2001-04-09</beginning_date>
    <ending_date>2001-07-06</ending_date>
  </dcterms:temporal>
  <platforms>
    <identifieur>mce.simu.all.sid.platforms</identifieur>
    <type>vv</type>
    <name>vv</name>
  </platforms>
  <memberlist>
    <identifieur>mce.simu.all.sid.platforms</identifieur>
  </memberlist>
  <workspaces>
    <identifieur>mce.simu.all.sid.workspaces</identifieur>
  </workspaces>
  <observations>
    <identifieur>mce.simu.all.sid.obs</identifieur>ce qui a ete observe durant
formation</observations>
  <research_instantiation>
    <identifieur>mce.simu.all.sid.research</identifieur>
  </research_instantiation>
</structured_interaction_data>
<rights_informed_consent>
  <head_part_corpus>
    <identifieur>mce.simu.all.ric.v1</identifieur>
    <description xml:lang="fr">xxx
  </description>
    <description xml:lang="en">xxx
  </description>
  </head_part_corpus>
</rights_informed_consent>
<analysis>
  <head_part_corpus>
    <identifieur>mce.simu.all.ric.v1</identifieur>
  </head_part_corpus>
</analysis>
</corpus>

```

### 12.3. Description LOM d'un scénario pédagogique

Contenu du fichier `mce_simu_all_ld_lom_v1.xml`, qui a été validé auprès du schéma LOM

```

<?xml version="1.0" encoding="utf-8"?>
<lom xmlns="http://www.imslobal.org/xsd/imsmd_v1p2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance">
  <general>
    <identifieur>mce.simu.all.rd.v1</identifieur>
    <title>
      <langstring xml:lang="en">Scenario for Simuligne</langstring>
    </title>
    <language>French as a foreign language</language>
    <description>
      <langstring xml:lang="en">XXX</langstring>
    </description>
    <description>
      <langstring xml:lang="fr">XXX</langstring>
    </description>
  </general>

```



```

</description>
<coverage>
  <langstring xml:lang="en">Simulation : pretexte scenario on summer language
university school ; Interculture : evryday situations in France and United
Kingdom</langstring>
</coverage>
<structure>
  <source>
    <langstring xml:lang="x-none">LOMv1.0</langstring>
  </source>
  <value>
    <langstring xml:lang="x-none">Hierarchical ; Networked</langstring>
  </value>
</structure>
<aggregationlevel>
  <source>
    <langstring xml:lang="x-none">LOMv1.0</langstring>
  </source>
  <value>
    <langstring xml:lang="x-none">4</langstring>
  </value>
</aggregationlevel>
</general>
<lifecycle>
  <version>
    <langstring xml:lang="x-none">1</langstring>
  </version>
  <status>
    <source>
      <langstring xml:lang="x-none">LOMv1.0</langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">Final</langstring>
    </value>
  </status>
  <contribute>
    <role>
      <source>
        <langstring xml:lang="x-none">LOMv1.0</langstring>
      </source>
      <value>
        <langstring xml:lang="x-none">Instructional Designer</langstring>
      </value>
    </role>
    <centity>
      <vcard>
BEGIN:VCARD
A.N:Hassan ; Xaviere
A.N:Lamy ; Marie-Noelle
ORG:The Open University
ADR:Milton Keynes ; United Kingdom
END:VCARD
BEGIN:VCARD
N:Chanier ; Thierry
ORG: Universite de Franche-Comte
ADR: Besancon ; France
END:VCARD</vcard>
      </centity>
    </contribute>
  </contribute>
  <role>
    <source>
      <langstring xml:lang="x-none">LOMv1.0</langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">Technical Implementer</langstring>
    </value>
  </role>
  <centity>
    <vcard>
BEGIN:VCARD

```

```

A.N: Reffay ; Christophe
A.N: Mbala ; Aloys
A.N: Dupont ; XX
A.N : XX ; Mohammed
ORG:Universite de Franche-Comte
ADR:Besancon ; France
END:VCARD
</vcard>
  </centity>
  </contribute>
</lifecycle>
<technical>mce.simu.all.sid.platforms</technical>
<educational>
  <interactivitytype>
    <source>
      <langstring xml:lang="x-none">LOMv1.0</langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">Active</langstring>
    </value>
  </interactivitytype>
  <learningresourcetype>
    <source>
      <langstring xml:lang="x-none">None</langstring>
    </source>
    <value>
      <langstring xml:lang="en">Web sites of cities ; questionnaires</langstring>
    </value>
  </learningresourcetype>
  <interactivitylevel>
    <source>
      <langstring xml:lang="x-none">LOMv1.0</langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">very high</langstring>
    </value>
  </interactivitylevel>
  <intendedenduserrole>
    <source>
      <langstring xml:lang="x-none">None</langstring>
    </source>
    <value>
      <langstring xml:lang="en">learner ; tutor ; native speaker</langstring>
    </value>
  </intendedenduserrole>
  <context>
    <source>
      <langstring xml:lang="x-none">LOMv1.0</langstring>
    </source>
    <value>
      <langstring xml:lang="x-none">higher education</langstring>
    </value>
  </context>
  <typicalagerange>
    <langstring xml:lang="en">adult</langstring>
  </typicalagerange>
  <difficulty>
    <source>
      <langstring xml:lang="x-none">None</langstring>
    </source>
    <value>
      <langstring xml:lang="en">Intermediate or advanced language level</langstring>
    </value>
  </difficulty>
  <typicallearningtime>
    <datetime>During 3 months</datetime>
  </typicallearningtime>
</educational>
<classification>
  <purpose>
    <source>

```

```

    <langstring xml:lang="x-none">LOMv1.0</langstring>
  </source>
  <value>
    <langstring xml:lang="x-none">Educational Objective</langstring>
  </value>
</purpose>
<taxonpath>
  <source>
    <langstring xml:lang="x-none">European Treasury Browser, ETB</langstring>
  </source>
  <taxon>
    <id>60 ; 70.20 ; 70.70 ; 50</id>
  <entry>
    <langstring xml:lang="en">distance education ; foreign language as teaching
language ; intercultural education ; project method</langstring>
  </entry>
</taxon>
</taxonpath>
<keyword>
  <langstring xml:lang="en">global simulation ; intercultural project ;
collaborative learning</langstring>
</keyword>
</classification>
</lom>

```

## 12.4. Entête fiche du corpus Simuligne avec espaces de noms

```

<mce:metadonnee xmlns:mce="http://mulce.univ-fcomte.fr/metadata"
xmlns:olac="http://www.language-archives.org/OLAC/1.0/"
xmlns:dc="http://purl.org/dc/elements/1.1/"
xmlns:dcterms="http://purl.org/dc/terms/"
xmlns:ims="http://www.imsglobal.org/xsd/imsmd_v1p2"
xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
xsi:schemaLocation="http://mulce.univ-fcomte.fr/metadata
mce_qualifieddc.xsd">
  <mce:header author="Thierry Chanier">
    <mce:identifiant>mce.simu.all.meta.v1</mce:identifiant>
    <mce:date>2007-11-27</mce:date>
    <mce:comment>La fiche de metadonnees du corpus est composee de 3
parties :
  - header : quelques informations sur cette fiche
  - olac : partie compatible Olac avec extension pour linguistic-type pour
"learning and teaching corpus"
  - corpus : partie propre a Mulce, inclut une partie Lom pour le scenario
pedagogique
    </mce:comment>
  </mce:header>

```

## 12.5. Annexe 1 : données Mulce venant d'autres sous-tâches

De Tdr1

### 12.5.1. Traitement et accessibilité des données

Ces paramètres s'applique aux données et au corpus (?)

- moderation : on\_line | off\_line
- access : anonymous | contractual | sensitive

### 12.5.2. Rôles

- Collecteur (de corpus) ;Éditeur (de corpus) ; Participant ; (Utilisateur) internaute ; (Utilisateur) internaute identifié ; (Utilisateur) contractuel ; Contributeur (de données) ; Responsable d'analyse

Seuls nous intéressent ici : Collecteur (de corpus) ;Éditeur (de corpus) ;Participant; Contributeur (de données) ; Responsable d'analyse

## 12.6. Annexe 2 : Descripteur corpus Clapi

[http://icar.univ-lyon2.fr/projets/corinte/integration/decrire\\_corpus-metadonnees.htm](http://icar.univ-lyon2.fr/projets/corinte/integration/decrire_corpus-metadonnees.htm)

Les descripteurs consultables en ligne sont :

#### Informations générales

Nom du corpus  
Date de recueil du corpus  
Corpus associés  
Durée globale du corpus  
Présentation du corpus  
Nombre de locuteurs du corpus  
Langue(s) des interactions

#### Responsable(s)

Nom  
Prénom  
Adresse mail

#### Auteur(s)

Nom  
Prénom  
Adresse mail

#### Bibliographie

Auteur(s) : nom, prénom  
Année  
Titre  
Type de publication  
Editeur  
Lieu de publication  
Page début  
Page fin  
Résumé  
Mots-clef

#### Enregistrement

Nom de l'enregistrement  
Présentation  
Support de l'enregistrement : Audio / Vidéo / Traces informatiques  
Mode de recueil : visible / caché / visible-caché / média  
Qualité de l'enregistrement : mauvaise / moyenne / bonne  
Est-il bipé, flouté ou anonymisé ? oui / non  
Durée de l'enregistrement  
Lieu(x) de recueil de l'enregistrement  
Nombre de locuteurs de l'enregistrement

Conditions de diffusion de l'enregistrement : Public / Communication  
et reproduction soumises à la signature d'une convention de recherche /  
Non diffusable

Accès à l'enregistrement

Documents annexes

Accès aux documents annexes (ex. scan de brouillons d'élèves)

Collecteur(s) de l'enregistrement

Nom

Prénom

Adresse mail

Genre interactionnel

Interactions collectées dans une situation construite par un  
chercheur : oui / non

Interactions médiatisés par ordinateur : oui / non

Interactions téléphoniques : oui / non

Interactions télévisuelles : oui / non

Interactions radiophoniques : oui / non

Interactions de service : oui / non

Interactions de commerce : oui / non

Interactions de travail : oui / non

Interactions en institution : oui / non

Interactions médicales : oui / non

Interactions didactiques : oui / non

Interactions privées : oui / non

Liste détaillée : Choix parmi la liste d'un ou plusieurs items ou  
ajout de nouveaux

ex. :

- co-rédaction d'un texte
- conférence
- cours de géographie en classe de cinquième
- consultation dentaire
- conversation dans un magasin de retouches
- conversation entre amis
- interactions en face à face sur un thème imposé
- négociation dans une étude de notaires
- ...

Transcription

Nom de la transcription

Support de la transcription :

- papier

- informatique : rtf, doc, pdf, txt, etc.

- alignée : Praat, CLAN, etc.

Orthographe : adaptée / standard

Transcription : totale / partielle

Transcription minutée : oui / non

Transcription anonymisée : oui / non

Transcription toilettée : oui / non

Transcription balisée : oui / non

Accès à la convention de transcription

Conditions de diffusion de la transcription : Public / Emprunt,  
Analyse et Requêtes soumis à la signature d'une convention de recherche /  
Analyses et Requêtes libres, Emprunt soumis à la signature d'une  
convention de recherche / Non diffusable

Accès à la transcription

Images annexes

Accès aux images annexes

Transcripteur(s)

Nom

Prénom

Adresse mail

Locuteurs

Nom ou pseudonyme du locuteur

Année de naissance

Age au moment de l'enregistrement

Lieu de naissance

Sexe

Situation au moment de l'enregistrement

Niveau d'études

Cursus suivi

Locuteur figurant également dans le corpus...../ l'enregistrement...

Langue(s) 1 / Langue(s) maternelle(s) du locuteur

Langue(s) 2 / Autre(s) langue(s)

Langue(s) de la mère

Langue(s) du père

Lieux de résidences successifs du locuteur

Professions successives du locuteur

Profession de la mère

Profession du père

Pays d'origine du locuteur

Pays d'origine de la mère

Pays d'origine du père

## 12.7. Note CR après discussion CR / TC du 5/11

### 12.7.1. Modification de Mulce-Struct :

Ancienne version de l'élément platform (qui était l'élément de plus haut niveau de l'instanciation) :

```
<xs:complexType name="platform">
  <xs:sequence>
    <xs:element ref="property"/>
    <xs:element ref="workspaces"/>
  </xs:sequence>
</xs:complexType>
```

Proposition pour la nouvelle structure (qui devient le nouvel élément de plus haut niveau de l'instanciation) qui se compose en 3 éléments comme suit :

```
<xs:complexType name="structured_interaction">
  <xs:sequence>
    <xs:element ref="platforms"/>
    <xs:element ref="memberlist"/>
    <xs:element ref="workspaces"/>
  </xs:sequence>
</xs:complexType>
```

où :

- \* platforms est une liste d'éléments "platform" (redéfinie ci-après)
- \* memberlist est la liste des acteurs participants à la situation d'apprentissage
- \* workspaces reste la liste des éléments workspace

Définition de platforms :

```
<xs:complexType name="platforms">
  <xs:element ref="platform" maxOccurs="unbounded"/>
</xs:complexType>
```

Redéfinition de platform (qui est en fait une reprise de l'ancien élément property de l'ex-platform, augmenté d'un champ):

```
<xs:complexType name="platform">
  <xs:sequence>
    <xs:element ref="type"/>
    <xs:element ref="name"/>
    <xs:element ref="version" minOccurs="0"/>
    <xs:element ref="url" minOccurs="0"/>
    <xs:element ref="designer" minOccurs="0"/>
    <xs:element ref="server" minOccurs="0"/>
  </xs:sequence>
</xs:complexType>

<xs:element name="type" type="xs:string"/>
```

type qui sera un champ de type text permettant de désigner l'environnement technologique de façon plus générique : LMS, Audio-Graphique Synchronous Environment, Blog, CMS, etc..

tandis que name reste plutôt le nom déposé (ou commercial) de la plateforme : WebCT, Lyceum, Blogger, etc...

### 12.7.2. Où et comment déclarer les contributeurs ?

Actuellement, nous avons choisi (dans la partie instanciation) de définir la liste des contributeurs dans l'élément `workspaces`. Nous gardons ce principe : c'est en effet à cet endroit que nous déclarons la liste des contributeurs, c.a.d. que **nous associons un code de référence à chaque contributeur (Prénom, Nom, Institution)**

En revanche, pour certains types de contributions (transcription, anonymisations, collecte, etc.) il est parfois utile d'être plus précis. En général, ces contributions sont réalisées sur les actes (associés à des outils) des `workspace` (exemple : les forums de simuligne, les actes de prod des sessions Tim1 à Tim8, les transcriptions audio de Copéas, etc.).

Nous **proposons comme principe d'attacher les contributions (contributeurs et rôles) à chaque outil dans chaque workspace**. Mais pour ne pas s'infliger de renseigner ces données dans chaque `workspace`, nous considérons que, en l'absence d'information sur les contributeurs d'un outil (ex de type forum) d'un `workspace`, nous remontons dans sa structure hiérarchique pour retrouver les contributeurs de l'outil de même type (ex : forum) dans le workspace englobant (ancêtre le plus proche) où cette information est spécifiée.

Ceci nous amène donc à modifier la structure de l'élément `tool` de la façon suivante :

```
<xs:complexType name="tool">
  <xs:sequence>
    <xs:element name="name" minOccurs="0">
      <xs:complexType>
        <xs:sequence>
          <xs:element name="designed" minOccurs="0"/>
          <xs:element name="collected" minOccurs="0"/>
        </xs:sequence>
      </xs:complexType>
    </xs:element>
  </xs:sequence>
```

```
    </xs:complexType>
  </xs:element>
  <xs:element name="accessibility" minOccurs="0">
    <xs:complexType>
      <xs:sequence>
        <xs:element ref="beginning_date"/>
        <xs:element ref="ending_date"/>
      </xs:sequence>
    </xs:complexType>
  </xs:element>
  <xs:element name="contribution" minOccurs="0" maxOccurs="1"/>
</xs:sequence>
<xs:attribute name="tool_type" use="required">
  <xs:simpleType>
    <xs:restriction base="xs:NMTOKEN">
      <xs:enumeration value="mail"/>
      <xs:enumeration value="chat"/>
      <xs:enumeration value="forum"/>
      <xs:enumeration value="dropbox"/>
    </xs:restriction>
  </xs:simpleType>
</xs:attribute>
<xs:attribute name="id" type="xs:ID" use="required"/>
</xs:complexType>
```

avec comme type pour contribution :

```
<xs:complexType name="contribution">
  <xs:sequence>
    <xs:element ref="contributor" maxOccurs="unbounded"/>
  </xs:sequence>
</xs:complexType>
```

et pour contributor :

```
<xs:complexType name="contributor">
  <xs:attribute name="idrefs" type="xs:IDS" use="required"/>
```

(syntaxe à vérifier pour dire qu'on déclare directement une liste d'identifiants dans l'ordre décroissant de l'importance de leur contribution sur ce "tool" et dans ce rôle)

```
  <xs:attribute name="role" type="xs:string" use="required"/>
</xs:complexType>
```

où le rôle serait pris dans la liste des rôles discuté dans métadonnées ci-avant..