



[U. Paris 3 Sorbonne nouvelle]

# mkMetadata

URL : <http://pi-ed268.univ-paris3.fr/>

Téléchargement : <http://pi-ed268.univ-paris3.fr/outils.html>

Version exécutable sous Windows disponible en ligne

Version script Perl/Tk disponible sur demande

(documentation en ligne sur la page du projet, ce document mis à jour)

# 1 Sommaire

## Version *OFFLINE*

1	Sommaire .....	2
2	Table des figures .....	2
3	Préambule.....	3
4	Présentation générale.....	4
5	Descriptif des icônes .....	5
6	Mode d'emploi .....	6
6.1	Ecriture des métadonnées.....	6
7	Export des traitements .....	8
8	Import de métadonnées .....	12

## Version *ONLINE*

9	Présentation générale.....	13
10	Génération et envoi des métadonnées .....	15
11	Références bibliographiques .....	16
11.1	Standards de codage de données linguistiques.....	16
11.2	Normes et standards concernant le codage des caractères .....	16
11.3	Normes et standards concernant les métadonnées .....	17

# 2 Table des figures

Figure 1	: MKM fenêtre principale .....	4
Figure 2	: MKM Onglet MKMETA1 .....	6
Figure 3	: MKM Onglet MKMETA3.....	6
Figure 4	: MKM Onglet MKMETA6.....	7
Figure 5	: MKM Editeur de métadonnées .....	8
Figure 6	: MKM Onglet EXPORT .....	9
Figure 7	: MKM Export HTML (1) .....	10
Figure 8	: MKM Export HTML (2) .....	10
Figure 9	: MKM Export PNG .....	11
Figure 10	: MKM online .....	14

### 3 Préambule

Ce programme est développé dans le cadre d'un projet intitulé *Proposition de Normalisation pour une Base de Corpus Multimédia à l'ED268* dont l'objectif est la création d'une base de données pluridisciplinaire, dans laquelle il sera possible d'archiver et de partager ses corpus oraux et vidéos. Ce projet innovant est financé par le Conseil Scientifique de l'Université Paris 3 - Sorbonne Nouvelle. Ce travail, réalisé dans un cadre pluridisciplinaire, réunit des chercheurs aux préoccupations scientifiques complémentaires, en confrontant les expérimentations diverses et les besoins émergents entre diverses disciplines utilisant des corpus de langue.

# mkMetadata offline

## 4 Présentation générale

Le programme **mkMetadata** (*offline*) permet de "construire" ou de corriger des métadonnées.

L'interface du programme est composée d'une fenêtre graphique disposant de différents onglets.

L'onglet **HOME** visible au chargement contient un texte d'identification du programme.

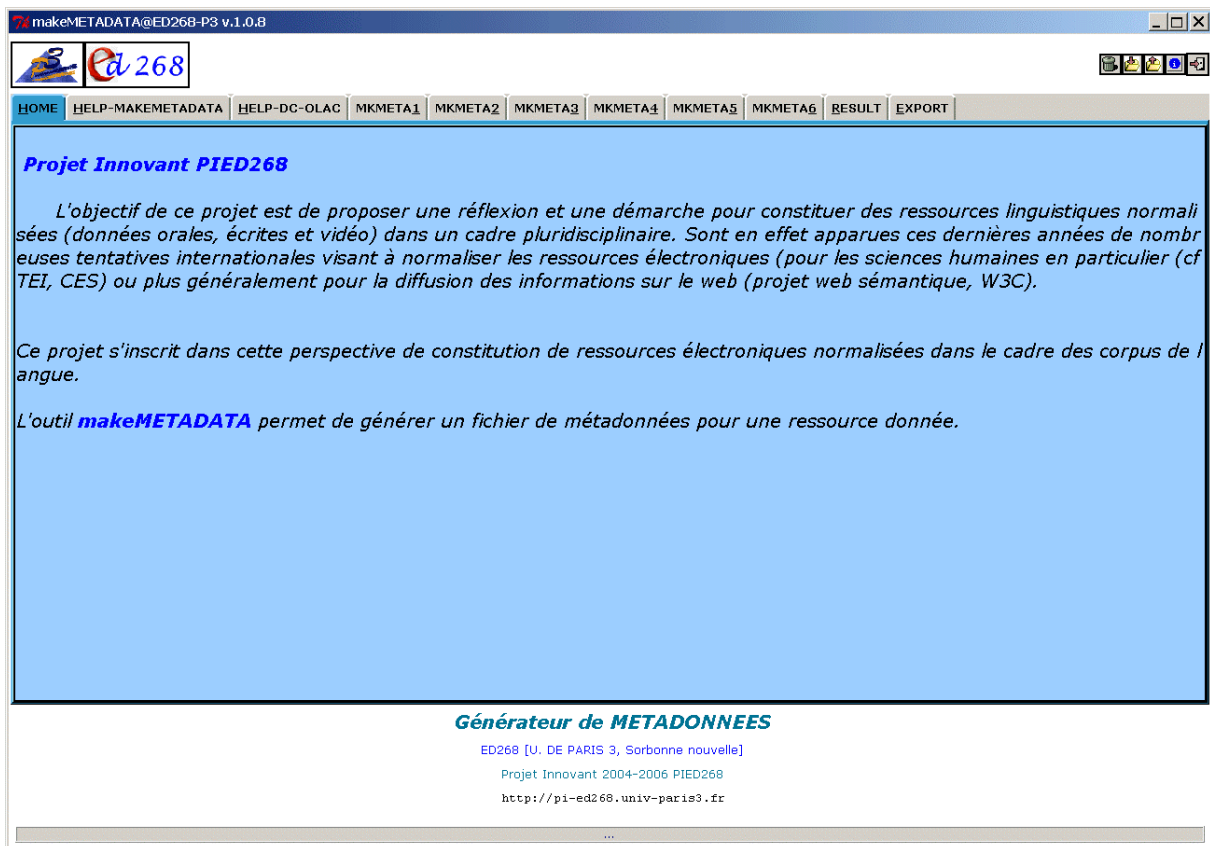



















Figure 1 : MKM fenêtre principale

Les autres onglets sont décrits *infra*.

## 5 Descriptif des icônes

<b>Icône</b>	<b>Fonction</b>	<b>Localisation</b>
	Sauvegarde (fichier source, cible, export, carte)	Onglets MKMETA, EXPORT
	Export au format HTML (feuille de style mkMetadata)	Onglet EXPORT
	Export au format PNG	Onglet EXPORT
	Export au format HTML (feuille de style OLAC)	Onglet EXPORT
	Parser XML	Onglet EXPORT
	Sortie des éditeurs de métadonnées	Editeur de Métadonnées
	Nouvel onglet dans les éditeurs de métadonnées	Editeur de Métadonnées
	Import d'alignement	Fenêtre principale
	Export au format XML	Fenêtre principale
	Documentation du programme	Fenêtre principale
	Réinitialisation des métadonnées	Fenêtre principale
	Sortie du programme	Fenêtre principale
	Accès au site d'OLAC	Onglet HELP-DC-OLAC
	Accès au codage des langues	Onglet HELP-DC-OLAC
	Accès au site Dublin Core	Onglet HELP-DC-OLAC
	Accès au site de l'ED	Fenêtre principale
	Accès au site de Paris 3	Fenêtre principale

## 6 Mode d'emploi

### 6.1 Ecriture des métadonnées

Les onglets **MKMETA1**, **MKMETA2**, **MKMETA3**, **MKMETA4**, **MKMETA5**, **MKMETA6** contiennent les fenêtres d'édition pour l'écriture des métadonnées et les points d'entrée pour les différentes fonctionnalités associées.

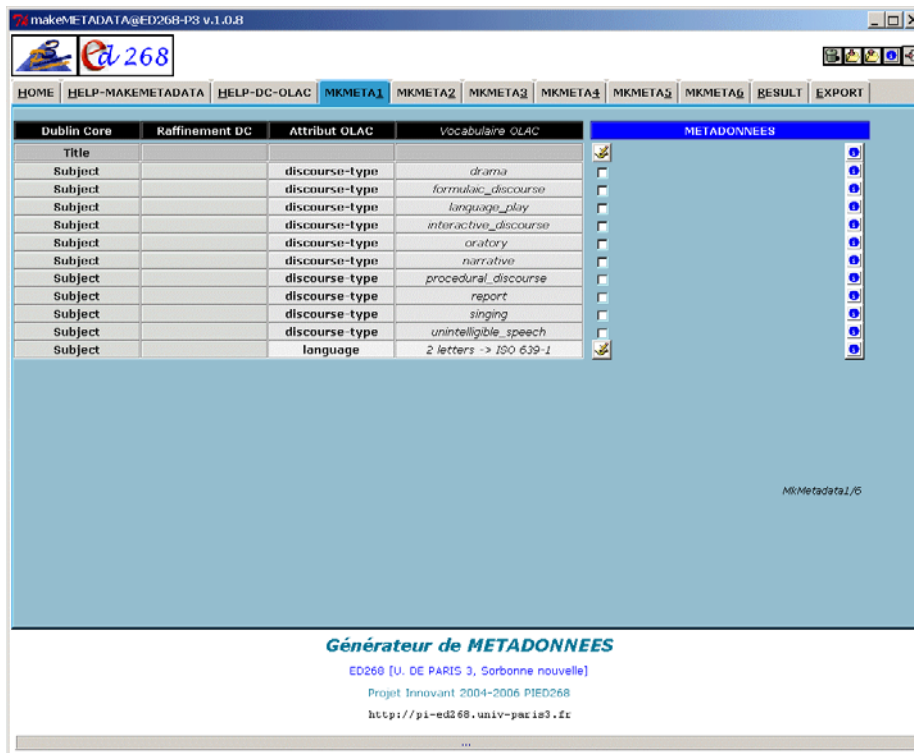


Figure 2 : MKM Onglet MKMETA1

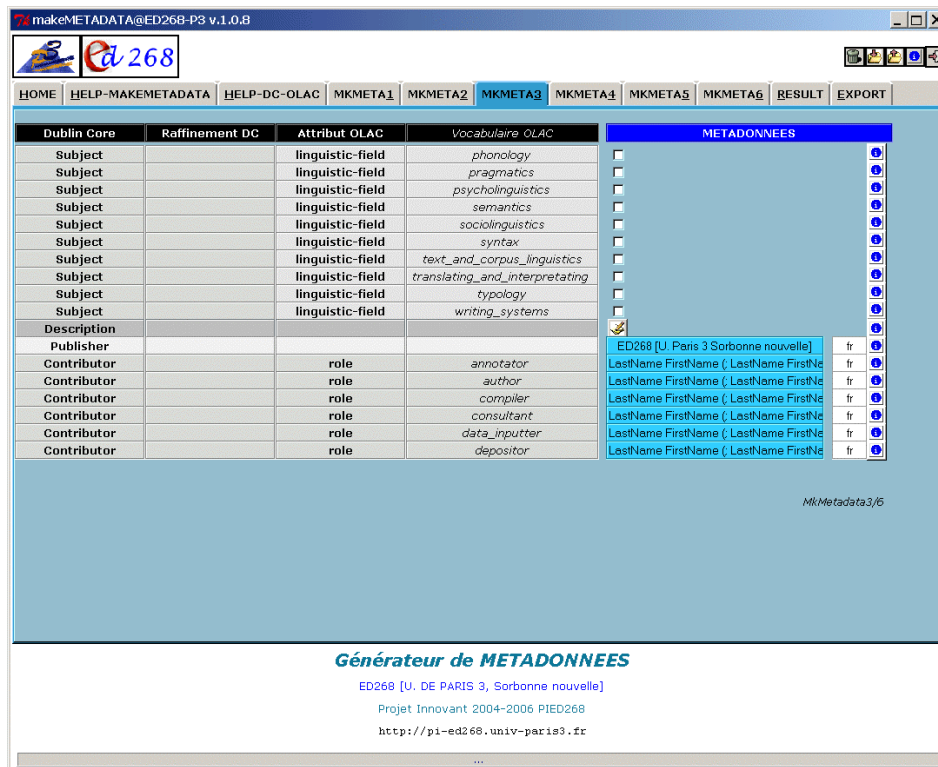


Figure 3 : MKM Onglet MKMETA3

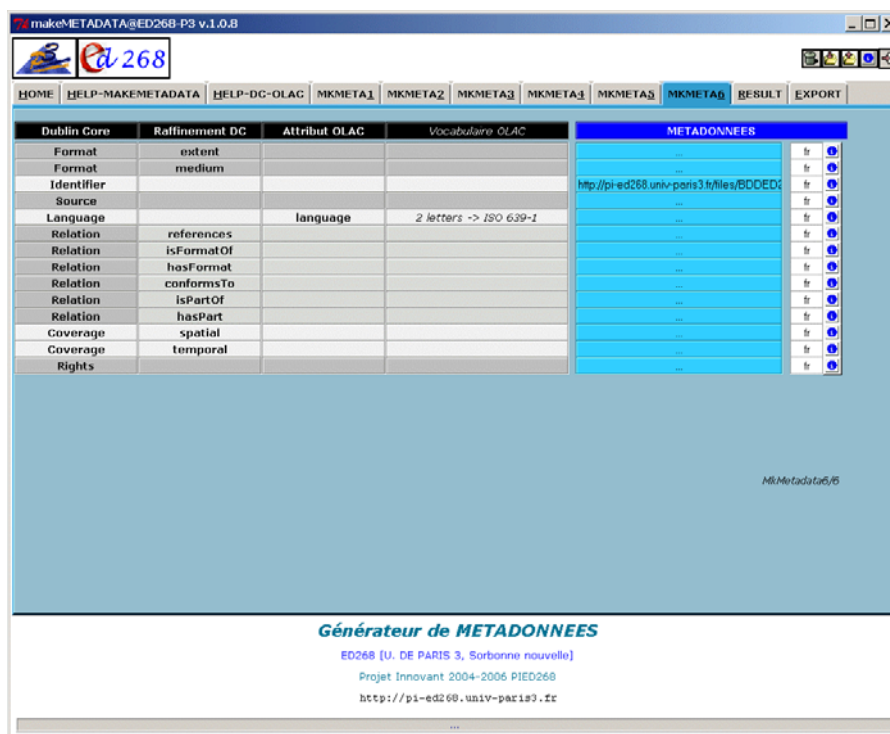


Figure 4 : MKM Onglet MKMETA6

Pour constituer les métadonnées, vous devez remplir l'ensemble des formulaires [MKMETA1](#), [MKMETA2](#), [MKMETA3](#), [MKMETA4](#), [MKMETA5](#), [MKMETA6](#).

Pour chacun de ces onglets, il faut compléter la colonne METADONNEES. Pour vous aider dans cette tâche vous pouvez consulter les fichiers d'aide disponibles (sur la colonne la plus à droite de chaque ligne du formulaire).

Dans chacun de ces onglets, vous trouverez soit des cases à cocher, soit des zones de saisie, soit un bouton "Edit" donnant accès à un *éditeur de métadonnées*.

Les zones de saisie se composent de deux champs (de saisie) :

- un pour entrer la valeur de la métadonnée idoine,
- l'autre pour décrire la langue utilisée dans le premier champ

Par défaut ce second champ est initialisé avec la valeur "fr".

Vous pouvez modifier cette valeur par défaut en utilisant le tableau de codage des langues utilisé par OLAC<sup>1</sup> et disponible à cette adresse :

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

vous pouvez aussi regarder le fichier HTML fourni avec **mkMetadata** (*LanguageCodes.html*) qui donne une présentation de ce tableau de codage des langues. Accès à ce fichier via le bouton **CODE** de l'onglet **HELP-DC-OLAC**.

Pour certaines zones de saisie, un modèle est fourni pour écrire la métadonnée : c'est le cas par exemple pour décrire les contributeurs : dans le cas où plusieurs personnes sont associées à un champ contributeur, chaque *Nom* complet (le nom suivi du prénom) doit être séparé du suivant par un point virgule. Il est important de respecter ce format.

<sup>1</sup> <http://www.language-archives.org/>

L'accès aux éditeurs de métadonnées permet de fournir une ou plusieurs descriptions pour une métadonnée donnée.

Pour chacune de ces descriptions, il est aussi possible de préciser la langue choisie pour écrire la métadonnée.

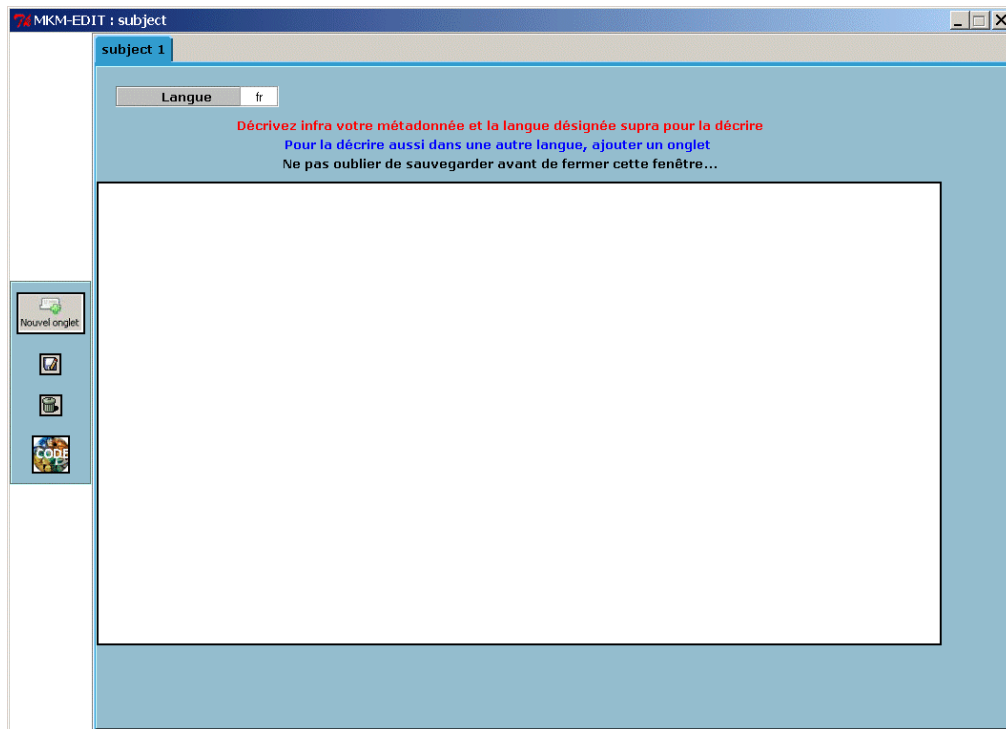


Figure 5 : MKM Editeur de métadonnées

### Exemple d'utilisation :

- on peut vouloir donner 2 métadonnées de type "title", l'une écrite en français et l'autre en chinois. Pour réaliser cette bi-description, on active l'éditeur de métadonnée disponible en regard de l'élément "title" 🛠️, une zone d'édition permet ensuite de décrire la métadonnée : dans celle-ci, un onglet = une description dans une langue donnée.

## 7 Export des traitements

Une fois les 6 onglets remplis, la génération des métadonnées est déclenchée par l'activation du bouton **EXPORT** 🛠️, le fichier de métadonnées constitué est visible dans la fenêtre d'édition de l'onglet **RESULT**.

La génération des métadonnées produit un fichier du type :

MK-METADATA-YYYYYYYYYYY.xml ,

où YYYYYYYYYY est une série de chiffres.

Ne pas oublier de renommer ce fichier après avoir quitté le programme, car au prochain lancement du programme tous les fichiers de ce type seront effacés (fichiers avec une extension html, txt, png etc.).



Au cours de la génération des métadonnées, une opération de transcodage de certains caractères est réalisée de la manière suivante :

- le caractère & est réécrit sous la forme de l'entité &amp; ;
- le caractère " est réécrit sous la forme de l'entité &quot; ;
- le caractère ' est réécrit sous la forme de l'entité &apos; ;
- le caractère < est réécrit sous la forme de l'entité &lt; ;
- le caractère > est réécrit sous la forme de l'entité &gt; ;

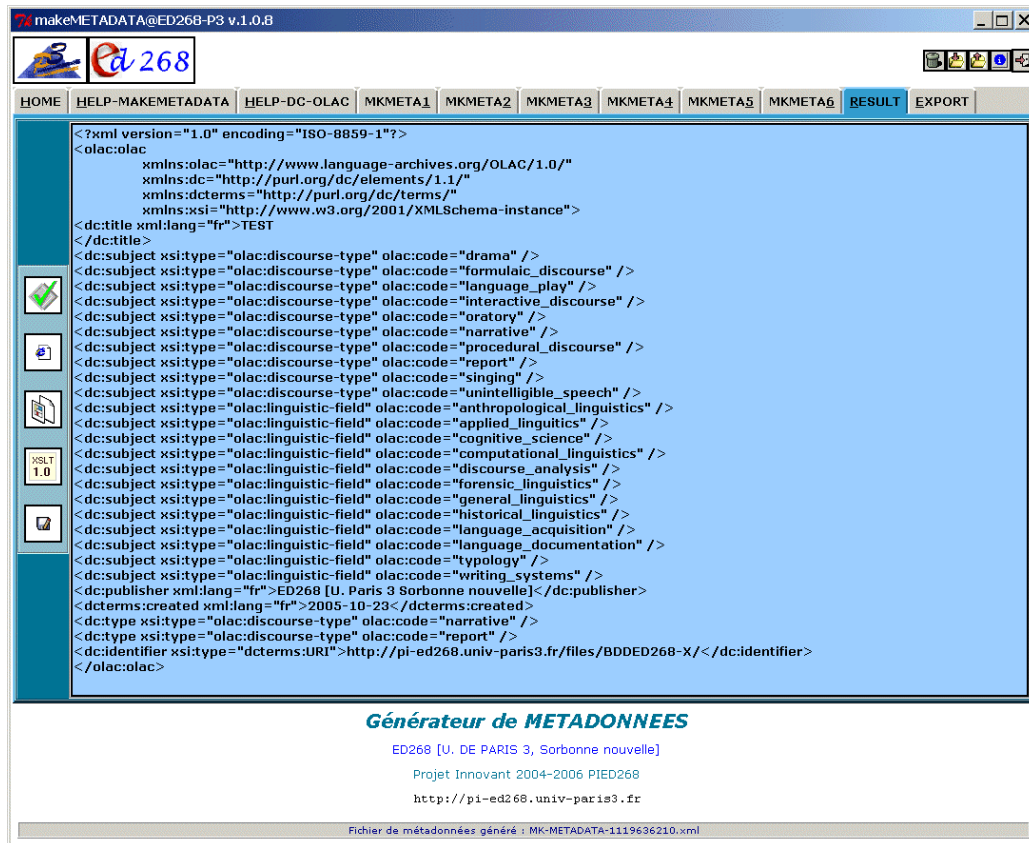


Figure 6 : MKM Onglet EXPORT

Une fois constitué le fichier de métadonnées, il est possible de valider ce document (parseur XML), de produire une version HTML de ce fichier de métadonnées (2 versions sont disponibles) ou d'en produire une image au format PNG.

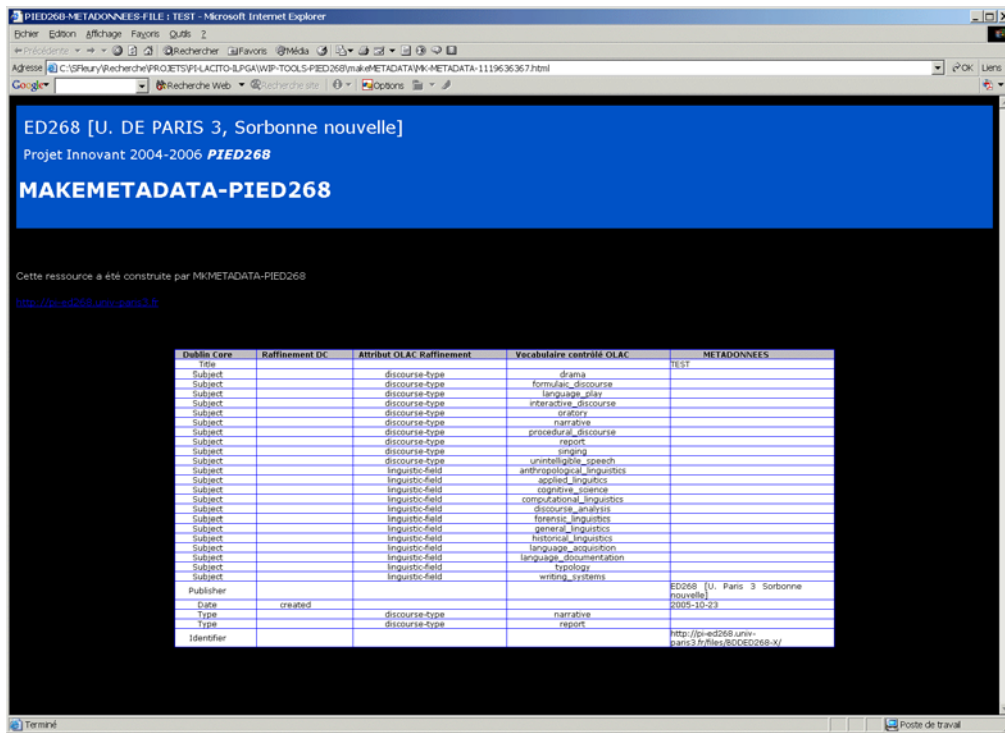


Figure 7 : MKM Export HTML (1)

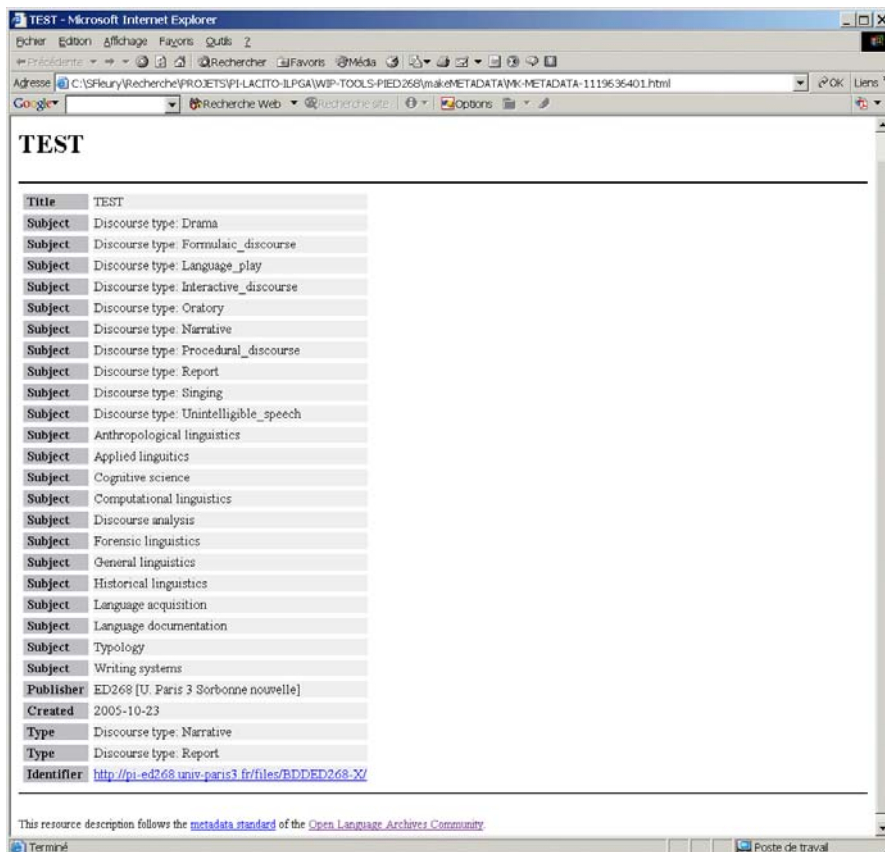


Figure 8 : MKM Export HTML (2)

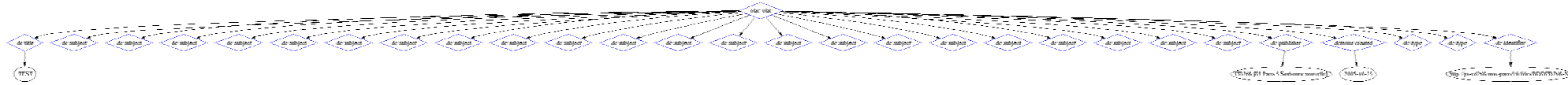




Figure 9 : MKM Export PNG

Le bouton **INIT**  (fenêtre principale) permet de réinitialiser l'ensemble des formulaires des onglets **MKMETA1**, **MKMETA2**, **MKMETA3**, **MKMETA4**, **MKMETA5**, **MKMETA6**.


## 8 Import de métadonnées

Il est possible de ré-importer un fichier de métadonnées préalablement construit avec ce programme. Au cours de la construction du fichier de métadonnées, ce programme construit un fichier d'export qu'il sera possible de ré-importer par la suite.

Un **fichier d'export** est automatiquement construit en activant le bouton **EXPORT** . Son nom est du type

EXPORT-MK-METADATA-YYYYYYYYYYY.xml

Ne pas oublier de renommer ce fichier après avoir quitté le programme, car à chaque lancement du programme tous les fichiers de ce type seront effacés.

Pour importer un fichier, il suffit d'activer le bouton **IMPORT**  et de sélectionner le fichier souhaité. Une fois le fichier sélectionné, l'import provoque la remise à jour des formulaires sur la base des items contenus dans le fichier importé.

# mkMetadata online

## 9 Présentation générale

Le programme **mkMetadata** (online) est une adaptation de la version offline pour le web.

Ce programme est disponible en ligne à cette adresse :

<http://pi-ed268.univ-paris3.fr/outils.html>

Après ouverture de la page du programme, la fenêtre du navigateur contient l'ensemble des fonctionnalités de **mkMetadata**.

On en présente une partie dans la figure suivante.

ED268 [U. DE PARIS 3, Sorbonne nouvelle]  
 Projet Innovant 2004-2006-PIED268  
**MAKEMETADATA-PIED268**

<http://u-ed268.univ-paris3.fr>

OLAC et DC  
 Dublin Core : <http://www.dublincore.org>  
 OLAC : <http://www.language-archives.org/>

Pour générer les métadonnées, modifier les champs ci-dessous. Les champs conservant les valeurs indéfinies par défaut "...\*" ou "LastName..." ne seront pas pris en compte dans les métadonnées, idem pour les cas où à cocher non activées.

Les zones de saisie se composent de deux champs (de saisie) : un pour entrer la valeur de la métadonnée idoine, l'autre pour décrire la langue utilisée dans le premier champ. Par défaut ce second champ est installé avec la valeur "fr". Vous pouvez modifier cette valeur par défaut en utilisant le tableau de codage des langues utilisé par OLAC et disponible à cette adresse : <http://www.language-archives.org/OLAC/OLAClanguagecodes.asp> ou vous pouvez regarder le fichier HTML fourni avec MakeMETADATA ici : <http://www.dublincore.org>

Pour certains éléments, il est possible de dupliquer les zones de saisie (clic sur bouton ) et si vous utilisez un navigateur de la famille Mozilla, vous pourrez éventuellement supprimer (clic sur ) les lignes insérées (cette suppression n'est pas possible avec Internet Explorer).

Dublin Core	Raffinement DC	Attribut OLAC	Raffinement	Vocabulaire contrôlé OLAC	MÉTADONNÉES	Code Langue	HELP
				+ [duplicate]			
Subject		discourse-type		drama	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		formulas_discourse	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		language_play	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		iterative_discourse	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		oratory	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		narrative	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		procedural_discourse	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		report	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		singing	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		discourse-type		unintelligible_speech	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		language		2 letters ->ISO 639-1		fr	HELP
				+ [duplicate]			
Subject		linguistic-field		anthropological_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		applied_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		cognitive_sciences	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		computational_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		discourse_analysis	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		forensic_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		general_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		historical_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		language_acquisition	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		language_documentation	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		lexicography	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		linguistics_and_literature	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		linguistic_theories	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		mathematical_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		morphology	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		neurolinguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		philosophy_of_languages	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		phonetics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		phonology	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		pragmatics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		psycholinguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		semantics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		sociolinguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		syntax	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		text_and_comput_linguistics	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		translating_and_interpreting	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		typology	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Subject		linguistic-field		writing_systems	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Description					Description de la ressource	fr	HELP
Contributor		role		research_participant	LastName FirstName [	fr	HELP
Contributor		role		responder	LastName FirstName [	fr	HELP
Contributor		role		signer	LastName FirstName [	fr	HELP
Contributor		role		singer	LastName FirstName [	fr	HELP
Contributor		role		speaker	LastName FirstName [	fr	HELP
Contributor		role		sponsor	LastName FirstName [	fr	HELP
Contributor		role		transcriber	LastName FirstName [	fr	HELP
Contributor		role		translator	LastName FirstName [	fr	HELP
Date	created				AAAA-MM-DD	fr	HELP
Date	valid				AAAA-MM-DD	fr	HELP
Date	available				AAAA-MM-DD	fr	HELP
Date	modified				AAAA-MM-DD	fr	HELP
Type		discourse-type		drama	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		formulas_discourse	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		language_play	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		iterative_discourse	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		oratory	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		narrative	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		procedural_discourse	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		report	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		singing	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		discourse-type		unintelligible_speech	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		linguistic-type		lexicon	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		linguistic-type		primary_text	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Type		linguistic-type		language_description	<input type="checkbox"/> NO <input type="checkbox"/> YES		HELP
Format	extent					fr	HELP
Format	medium					fr	HELP
Identifier					<a href="http://pied268.univ-paris3.fr">http://pied268.univ-paris3.fr</a>	fr	HELP
Source						fr	HELP
Language		language		2 letters ->ISO 639-1		fr	HELP
Relation	references					fr	HELP
Relation	isFormatOf					fr	HELP
Relation	hasFormat					fr	HELP
Relation	conformsTo					fr	HELP
Relation	isPartOf					fr	HELP
Relation	hasPart					fr	HELP
Coverage	spatial					fr	HELP
Coverage	temporal					fr	HELP
Rights						fr	HELP

Votre Nom : [Nom Prénom]

Figure 10 : MKM online

## 10 Génération et envoi des métadonnées

Après avoir complété l'ensemble des champs du formulaire, l'ensemble des données est transmis par mail aux responsables du projet

## 11 Références bibliographiques

- Bird Steven and Gary Simons (2004), "Building an Open Language Archives Community on the DC Foundation", in Hillmann and Westbrook (editors), Metadata in Practice : A Work in Progress, ALA Editions
- Bird Steven, M. L. (2000). "A formal Framework for Linguistic annotation." Speech Communication **33** ((1,2)): 23-60.
- Habert Benoît, F. C. e. I. F. (1998). De l'écrit au numérique : constituer, normaliser, exploiter les corpus électroniques. Paris, InterÉditions/Masson.
- Habert Benoît, N. A. e. S. A. (1997). Les linguistiques de corpus Paris, Armand Colin/Masson.
- Harold E.R., W. S. M. (2001). XML in a nutshell, O'REILLY.
- IDE, N., VÉRONIS, J. (1994). MULTEXT (Multilingual Tools and Corpora). Proceedings of the 14th International Conference on Computational Linguistics, COLING'94, Kyoto, Japan.
- IDE Nancy, J. V. (1996). "Une application de la TEI aux industries de la langue : le Corpus Encoding Standard." Cahiers GUTenberg **24**.
- IDE Nancy, V. J. (1996). "Présentation de la TEI : Text Encoding Initiative." Cahier Gutenberg **24**: 4-10.
- Sandoval, V. (1996). SGML, HERMES.
- Véronis, J. (2000). Annotation automatique de corpus : panorama et état de la technique. Ingénierie de langues. J. M. Pierrel. Paris, Hermès.

### 11.1 Standards de codage de données linguistiques

- L'alphabet de phonétique internationale : <http://www.arts.gla.ac.uk/ipa/ipa.html>
- La TEI "text encoding initiative" : <http://www.tei-c.org>
- Le corpus encoding standart : <http://www.cs.vassar.edu/CES/>
- Le groupe ISO : <http://www.tc37sc4.org/> en dehors du site il doit y avoir de la littérature par Laurent Romary et/ou Nancy Ide.

### 11.2 Normes et standards concernant le codage des caractères

- Le site officiel du consortium Unicode: <http://www.unicode.org>
- André, J., " Caractères, codage et normalisation - de Chappe à Unicode ", Document numérique, vol. 6, n° 3-4, 2002, pages 13 à 49. ([http://www.irisa.fr/fagtypo/unicode/chappe\\_a\\_unicode.pdf](http://www.irisa.fr/fagtypo/unicode/chappe_a_unicode.pdf))
- André, J. et Goosens, M., " Codage des caractères et multi-linguisme: de l'ASCII à Unicode et ISO/IEC-10646 ", Les cahiers de GUTenberg, n°20, mai 1995. (<http://www.gutenberg.eu.org/pub/GUTenberg/publicationsPDF/20-jamg.pdf>)
- Andries, P., " Entretien avec Ken Whistler, directeur technique du consortium Unicode ", Document numérique, vol. 6, n° 3-4, 2002. (<http://pages.infinet.net/hapax/pdf/whistler.pdf>)
- Andries, P., " Introduction à Unicode et à l'ISO 10646 ", Document numérique, vol 6, n°3-4, 2002, pages 51-88. (<http://iquebec.ifrance.com/hapax/pdf/intro-Unicode.pdf>)
- Benoit, J.-L., Bernet, Ch, Bonhomme, P., Romary, L. et Viscogliosi N, " Du document électronique à son usage : le rôle central de la normalisation ", Solaris, n°6, janvier 2000 (numéro complet : <http://biblio-fr.info.unicaen.fr/bnum/jelec/Solaris/d06/index.html>)



### *11.3 Normes et standards concernant les métadonnées*

- Norme de métadonnées pour le codage des ressources informatiques: <http://dublincore.org/>
- Standard de codage de métadonnées utilisé par les bibliothèques: <http://lcweb.loc.gov/marc/>
- Organisation d'archivistes de données linguistiques et proposition de codage de métadonnées pour cette communauté: <http://www.language-archives.org/>
- IMDI (ISLE Meta Data Initiative) autre proposition pour la communauté linguistique : <http://www.mpi.nl/IMDI/>
- Standard de diffusion de métadonnées: <http://www.openarchives.org/>