

## [mdf-italie] Compte rendu de la réunion du 19 janvier

**Objet** : premières mises en ligne pour l'évaluation de l'ANR fin mars 2015

Participants	1
Projet éditorial	1
Catalogue des articles	1
Articles transcrits	2
Modèle de métadonnées jusqu'en mars 2015	2
Exemples	4
Discussion des métadonnées	7

### Participants

Davide Luglio	davide.luglio@paris-sorbonne.fr
Frédéric Glorieux	frederic.glorieux@paris-sorbonne.fr
Eric Thiébaud	thiebaud.eric01@gmail.com
Iris Berger	iris.berger.iris@gmail.com
Alexandra Ivanovitch	alexandra.ivanovitch@gmail.com

### Projet éditorial

Avant l'OBVIL, le projet comportait deux volets : un catalogue qui recense les articles du *Mercure de France* (1890-1918) concernant la culture italienne, et l'édition d'une anthologie d'articles importants. Le catalogue est en partie établi, l'anthologie n'a pas été commencée. Au sein de l'OBVIL, le projet consiste à numériser et transcrire tous les articles recensés par le catalogue.

### Catalogue des articles

Les articles du *MdF* concernant la culture italienne ont été recensés par l'équipe scientifique sous la forme d'un catalogue en traitement de texte. Au moins pour les années 1903-1918, chaque article est décrit avec les informations suivantes :

- Auteur(s) ;
- Rubrique, ou titre d'article ;
- Localisation bibliographique ;
- Résumé.

La totalité de ce catalogue peut être mise en base de données d'ici mars 2015, afin de présenter le périmètre du projet. Toutefois, les résumés ne seront pas tout de suite exposés. Outre une question de droits, il faut aussi tester leur apport relativement au texte des articles.

Les tâches à réaliser :

- stylage et normalisation du traitement de textes ;
- traitement, correction, et validation du XML (notamment les identifiants) ;
- insertion dans une base de données SQL ;
- développement d'une application de consultation en ligne.

### **Articles transcrits**

Les articles transcrits seront à terme insérés dans une base de données pour exploiter l'indexation et permettre la recherche plein texte. Toutefois, dans l'état de la réflexion, nous avons examiné la possibilité de produire un objet éditorial avec une compilation d'articles sur une année, organisés par rubrique du *MdF*, et rangés par dates. Ce classement humain permettrait de vérifier les titres et d'assurer que les rubriques sont pertinentes pour distinguer les types de contenus. Ce type de compilations ne se substitue pas aux anthologies, car elles sont exhaustives et sans ajout critique. Cette éditorialisation aurait cependant l'intérêt de présenter les articles dans une série « naturelle », en évitant le caractère intimidant d'un formulaire de recherche vide. Ces séries organisées feront l'objet d'une génération automatique de livres électroniques pour lecture hors ligne. Le travail à réaliser consisterait alors à :

- Numériser le texte des revues papier (OCR, Abbyy)
- Corriger, organiser, styler (Titre 1 : rubrique, Titre 2 : titre forgé d'un article), et ajouter les métadonnées aux articles en un fichier traitement de texte
- Convertir en XML/TEI et valider les métadonnées
- Présenter en ligne les compilations
- Relier les articles à la base du catalogue (mars 2015)

### **Modèle de métadonnées jusqu'en mars 2015**

Chaque article extrait sera muni de plusieurs champs de métadonnées. La première campagne d'indexation se concentrera sur les informations qui ne nécessitent pas une connaissance globale du corpus ; aussi cette section n'abordera-t-elle pas l'indexation thématique, sur laquelle on reviendra en fin de document.

Chaque champ est identifié par un mot-clé et par un quantifieur. Les quantifieurs sont indiqués selon la notation des expressions régulières. Cette

précision est importante dès maintenant, car elle a des conséquences sur la modélisation et le traitement informatique du projet.

?	0-1	optionnel
!	1	obligatoire
+	1-n	obligatoire et répétable
*	0-n	optionnel et répétable

**(! title)** : chaque article doit être titré, dans le contexte hiérarchique de sa compilation annuelle (table des matières). Les articles titrés conservent leur titre, un titre sera forgé pour les autres articles (ex : extrait de rubrique). Placé entre crochets, le titre forgé sera explicite et synthétique.

**! identifiant** : mdf\_année-mois-jour\_page, ex. mdf\_1903-07-01\_262. Chaque article doit être identifié avec un code forgé. Cet identifiant doit être unique dans le contexte du Labex, et être compatible avec des projets futurs autour du *Mercure de France*. Le tiret bas « \_ » est un séparateur d'un niveau supérieur à celui du trait d'union « - ». L'identifiant combine trois membres : 1) un préfixe pour le projet, 2) une date, 3) un numéro de page. Le préfixe est en minuscules non accentuées : mdf. Une date suffit à identifier chaque numéro de la revue (mensuelle, puis bimensuelle à partir de 1905) ; elle est en chiffres, du générique au spécifique : AAAA-MM-JJJJ (ex : 1912-01-16). La page de début suffit généralement à identifier chaque article. Si plusieurs articles courts commencent sur une même page, ajouter une lettre, ex : 262a (premier article de la page 262), 262b (deuxième article de la page 262).

**! date** : Chaque article doit être daté au format AAAA-MM-JJ (année-mois-jour). On jugera en cours d'exercice s'il vaut mieux retrouver automatiquement cette date depuis l'identifiant.

**! source** : chaque article doit être décrit par une référence bibliographique ponctuée et typographiée (texte riche), il suffit généralement de reprendre la référence déjà composée dans le catalogue.

**! rubric** : on conservera toujours l'intitulé original de la rubrique *MdF* de l'article, ce qui autorise la reconstitution des compilations annuelles. Information déjà présente dans le catalogue.

**+ creator** : un article a toujours au moins un auteur. Pour les articles non signés, souvent courts, le catalogue a adopté la convention : [Rédaction]. Il est inutile de détailler l'équipe. Pour les auteurs multiples, répéter le champ *creator*. Les noms d'auteur seront inscrits sous une forme d'autorité de type BNF : Nom, Prénom (année de naissance-année de mort). En cas de pseudonyme : Pseudonyme [Nom, Prénom] (année de naissance-année de mort). Source principale de normalisation : <http://catalogue.bnf.fr>. Autres sources consultées : <http://opac.bncf.firenze.sbn.it> ; <http://www.idref.fr/autorites/autorites.html>

? **translator** : un texte peut avoir été traduit, indiquer le traducteur selon les mêmes normes qu'un auteur.

\* **bibl** : zéro ou plusieurs œuvres (ou seulement auteur) sujet de l'article. Ce champ *œuvre* est à entendre au sens large de « biblio/média-graphique », étant donné la portée multimédiatique du projet de recherche. Seules les œuvres qui font l'objet d'un développement assez nourri seront mentionnées ici. La constitution de ce champ pourra être faite à partir des résumés du catalogue, où est déjà effectué le travail d'extraction et de pondération. Pour être exploitable, la référence sera structurée du générique au spécifique : « Auteur(s). Titre. ... » Le point « . » est le séparateur des membres bibliographiques, notamment entre le groupe « auteurs » et le groupe « titres ». L'application exploitera principalement le regroupement alphabétique par auteur, et par titre. Le contenu après le deuxième point est libre. Le point-virgule est le séparateur entre les auteurs « ; ».

- Article consacré à un auteur : Nom, Prénom (année de naissance-année de mort)

- Recension d'une œuvre imprimée (parution) : Nom, Prénom (année de naissance-année de mort). Titre : sous-titre. Ville, éditeur, année de parution, volumes.

- Recension d'une œuvre vivante (représentation) : Nom, Prénom (année de naissance-année de mort). Titre : sous-titre. Représentations

### Exemples

Chaque exemple reprend un enregistrement du catalogue et présente la manière dont il peut apparaître dans un cartouche de métadonnées, à placer sous le titre de l'article.

ZUCCOLI LUCIANO

*Lettres italiennes*

Juillet 1903, T. XLVII, n. 163, p. 262-266. **5/67**

La chronique consacre une partie différente aux trois domaines de la production littéraire : le théâtre, le roman et la poésie. Pour le théâtre, Z. cite : E.A. Butti, *Il gigante e i Pigmei*, un drame passionnel qui se passe dans un milieu littéraire. Les Italiens ont d'ailleurs reconnu dans ses personnages des écrivains contemporains, célèbres et moins célèbres (de Carducci à Luciano Zuccoli lui-même !). Après ce fait divers, Z. rappelle rapidement le triomphe au théâtre de Renato Simoni (*Vedova*), de Roberto Bracco (*Maternità*), de G. Antona Traversi (*I giorni più lieti*). Marco Praga avec *Ondina* n'a pas eu le même succès. Parmi les romans dernièrement édités, Z. rappelle seulement : *Sfinge*, *Dopo la vittoria* (Milano, Treves), un roman dont il juge, d'ailleurs, l'intrigue assez faible. La partie de la chronique consacrée à la poésie est partagée entre deux poètes : D'Annunzio et Pascoli. Du premier, Z. cite l'édition des *Laudi* (Milano, Treves, avec des dessins de Giuseppe

Cellini) et fait allusion au scandale que ce recueil a créé dans le milieu clérical à cause de son paganisme. Annonçant l'édition des *Canti di Castelvechio*, Z. affirme que, dans ce recueil, Pascoli montre avoir su exploiter « cette vie naïve, il goûte les rythmes populaires, il étudie les refrains, il imite le gazouillage des oiseaux, et souvent, d'un coup, par les choses et les sentiments simples de la terre, il arrive à des méditations et à des comparaisons profondes » (p. 266). Par sa vision profonde de l'univers, Pascoli est comparé à Shelley.

identifiant : mdf\_1903-07-01\_262  
 date : 1903-07-01  
 source : Tome XLVII, n° 163, p. 262-266  
 rubric : Lettres italiennes  
 creator : Zùccoli, Luciano (1868-1929)  
 bibl : Butti, Enrico Annibale (1868-1912). *Il Gigante e i Pigmei : comedia in quattro atti*. Représentations  
 bibl : Simoni, Renato (1875-1952). *La Vedova : comedia in tre atti*. Représentations  
 bibl : Bracco, Roberto (1862-1943). *Maternità : dramma in quattro atti*. Représentations  
 bibl : Antona Traversi, Giannino (1860-1939). *I giorni più lieti : comedia in tre atti*. Représentations  
 bibl : Praga, Marco (1862-1929). *L'Ondina : dramma in quattro atti*. Représentations  
 bibl : Sfinge [Condrochi Argeli, Eugenia] (1865-1934). *Dopo la Vittoria : Romanzo*. Milano : Fratelli Treves, 1903.  
 bibl : D'Annunzio, Gabriele (1863-1938). *Laudi del cielo, del mare, della terra e degli eroi*. Milano : Fratelli Treves, 1903-1904, 2 vol.  
 bibl : Pascoli, Giovanni (1855-1912). *Canti di Castelvechio*. Bologna : Zanichelli, 1903.

SYMONS ARTHUR (trad. par DAVRAY HENRY D.)  
 Casanova à Dux. Un chapitre d'histoire inédit  
 Octobre 1903, T. XLVIII, n. 166, p. 60 88. 29/110  
 Étude consacré aux Mémoires de Casanova. Symons affirme en avoir vu à Leipzig le manuscrit et rend compte des parties omises dans l'édition Brockhaus (1822 1828, traduction allemande).

identifiant : mdf\_1903-10-01\_60  
 date : 1903-10-01  
 source : Tome XLVIII, n° 166, p. 60  
 creator : Symons, Arthur (1865-1945)  
 translator : Davray, Henry D. (1873-1944)  
 bibl : Casanova, Giacomo (1725-1798). *Histoire de ma vie*

MORICE CHARLES

Art moderne

Décembre 1903, T. XLVIII, n. 168, p. 817-822 [820]. **6/167 1903 = 167**

Annonce d'une exposition Zandomenighi (aux Galeries Durand Ruel). M. juge le peintre italien peu original : « Tout est bleu et rose chez M. Zandomenighi [...] Tout est doux, lisse et rond » (p. 820). Néanmoins, il apprécie l'ensemble, qui est toujours mélodieux.

identifier : mdf\_1903-12-01\_820

date : 1903-12-01

source : Tome XLVII, no 168, p. 820

rubric : Art moderne

creator : Morice, Charles (1860-1919)

bibl : Zandomenighi, Federico (1841-1917). *Exposition Galerie Durand-Ruel (Paris)*

MARNOLD JEAN

*Musique*

Avril 1904, T. L, n. 172, p. 242-251 [242-248]. **10/35**

Une partie de la chronique consacrée aux théâtres d'opéra parisiens rend compte du succès de la représentation de *L'Orfeo* de Monteverdi, « une œuvre inconnue autant qu'illustre [qui] vient d'être exhumée de la poussière des bibliothèques » (p. 242). *Orfeo* a été reconstitué d'une manière assez fidèle, mais en tenant compte des moyens dont dispose la scène contemporaine. Surtout, selon M., l'intérêt de représenter *l'Orfeo* se situe dans ce fait qu'il ressuscite une phase de l'évolution musicale analogue à celle du dernier siècle (celle de l'harmonie, soit la phase où la musique est la « servante » du drame). M. affirme que Monteverdi, par son art primitif, peut être rapproché de Debussy, alors que pour son *stile rappresentativo* (déclamation dramatique) il a été rapproché de Wagner.

identifier : mdf\_1904-04-01\_242

date : 1903-04-01

source : Tome L, n° 172, p. 242-248

rubric : Musique

creator : Marnold, Jean (1859-1935)

bibl : Monteverdi, Claudio (1567-1643). *L'Orfeo*. Représentations

Davray Henry D.

*Lettres anglaises*

Juin 1904, T. L, n. 174, p. 819-827 [821-822]. **8/56**

Compte rendu de l'ouvrage de Arthur C. Benson, *Rossetti* (éd. Macmillan, collection *English Men of Letters*). Il s'agit d'une étude sur Dante Gabriele Rossetti poète et peintre préraphaélite, mais aussi sur l'influence qu'il a exercée dans la culture anglaise. Elle retrace de même l'histoire de la famille du réfugié italien à Londres.

identifier : mdf\_1904-06-01\_821  
date : 1904-06-01  
source : Tome L, n° 174, p. 821-822  
rubric : Lettres anglaises  
creator : Davray, Henry D. (1873-1944)  
bibl : Benson, Arthur Christopher (1862-1925). *Rossetti (Dante Gabriel)*.  
London, Macmillan, 1904  
bibl. : Rossetti, Dante Gabriel (1828-1882)

### Discussion des métadonnées

Les métadonnées ci-dessous restent très factuelles et ne permettent pas, par exemple, de trouver tous les articles longs qui concernent la peinture. Il a paru important d'attendre d'avoir une masse numérisée suffisante avant d'alourdir ce premier programme d'indexation. En croisant différentes approches, on gardera ouvert le champ des possibles interrogations et exploitations du corpus à venir.

La solution envisagée présuppose donc d'allier deux logiques documentaires complémentaires et de combiner :

- une approche centrée sur le type documentaire, avec la formalisation de données fidèlement reprises à la matrice éditoriale originale (cf les rubriques du *MdF*), ainsi qu'une orientation objet avec l'extraction et la normalisation des références biblio-médiagraphiques des œuvres étudiées ou recensées dans le *MdF* ;
- une infrastructure sujet, avec un système de mots-clés qui facilitera les recherches et regroupements de textes par thème, par média traité, etc., pour d'éventuelles publications à venir.

Un tel compromis permettrait à la fois des recherches extrêmement ciblées, portant sur des objets précis, et des recherches générales, se fondant sur des critères plus larges.

On procèdera à des tests sur un corpus circonscrit pour mettre à l'épreuve la pertinence de la liste de mots-clés retenus.

Certains « thèmes » (par ex. : futurisme) peuvent tirer parti d'une approche externe, où l'indexation ne se fait pas au moment de l'informatisation d'un article, mais sous la forme d'une liste de liens (pointant vers des articles), une fois le corpus mis en ligne. Cette manière permet de garantir l'exhaustivité de la liste proposée sur un sujet, d'autant plus qu'il vaut mieux laisser l'appréciation du sujet à des chercheurs spécialistes des domaines envisagés.