



**HAL**  
open science

## Le programme de recherche sur les archives du Fonds Alcides Giraldi : un exemple de production de données en humanités numériques

Cécile Braillon-Chantraine, André Davignon, Nicolas Lahoche, Fatiha Idmhand, Cécile Martini

### ► To cite this version:

Cécile Braillon-Chantraine, André Davignon, Nicolas Lahoche, Fatiha Idmhand, Cécile Martini. Le programme de recherche sur les archives du Fonds Alcides Giraldi : un exemple de production de données en humanités numériques. *Revue des Nouvelles Technologies de l'Information*, 2014, *Fouille de Données et Humanités Numériques*, numéro spécial de la *Revue des Nouvelles Technologies de l'Information*, p.169-184. halshs-01132614

**HAL Id: halshs-01132614**

**<https://shs.hal.science/halshs-01132614>**

Submitted on 17 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Le programme de recherche sur les archives du Fonds Alcides Giraldi : un exemple de production de données en humanités numériques<sup>1</sup>

Cécile Braillon-Chantraine\*, André Davignon\*\*, Nicolas Lahoche\*\*\*,  
Fatiha Idmhand\*\*\*\*, Cécile Martini<sup>2</sup>\*\*\*\*\*

\*Institut des Textes et Manuscrits Modernes (ITEM, UMR 8132)  
École Normale Supérieure  
45, rue d'Ulm 75005 Paris  
fatihaidmhand@yahoo.es  
<http://www.item.ens.fr>

\*\*Centre de recherches sur les civilisations, les langues et les littératures étrangères  
(CECILLE, EA4074)  
Université des Sciences Humaines et Sociales - Lille3  
Domaine universitaire - Pont de Bois  
59653 Villeneuve d'Ascq Cedex  
cecile.chantraine@sfr.fr  
<http://cecille.recherche.univ-lille3.fr/>

Article publié dans *Fouille de Données et Humanités Numériques*, numéro spécial de la Revue des Nouvelles Technologies de l'Information, RNTI SHS-2, Sous la direction de Djamel A. Zighed et Gilles Venturini, Rédacteur invité : Nicolas Turenne (INRA-SenS - UPEM), p.169-184, 2014.

---

<sup>1</sup> Mots-clés : Données SHS, Fouille de connaissances, Fouille du web, Modèles d'archivage et de création de corpus, Plateformes numériques, Qualité et standardisation des données, Ressources numériques, ontologies et métadonnées, Standards numériques pour la gestion des données, Visualisation des connaissances.

<sup>2</sup> Cécile Braillon-Chantraine est Maître de Conférences en Lettres et Langues (espagnol) à l'Université de Valenciennes (UVHC, Lille Nord de France) et rattachée au centre de recherches Centre d'Études et de Recherches sur Civilisations, les Langues et les Littératures Étrangères (CECILLE EA4074, Lille 3) ; André Davignon est chef de projet et responsable du Point d'Appui National Documentaire (PANDOC) au ministère de l'écologie, du développement durable et de l'énergie (PANDOC, Lille), Nicolas Lahoche est webmestre et responsable de la communication web à l'Université Lille 3, Fatiha Idmhand est Maître de Conférences à l'Université du Littoral en Lettres et Langues (espagnol) à L'Université du Littoral Côte d'Opale (ULCO, Lille Nord de France) et rattachée au centre de recherches Centre d'Études et de Recherches sur Civilisations, les Langues et les Littératures Étrangères (CECILLE EA4074, Lille 3) et Cécile Martini est archiviste et paléographe, directrice adjointe du Service Commun de la Documentation de l'Université Lille 3 (Lille Nord de France).

**Résumé.** Parmi les patrimoines écrits que les institutions publiques ou privées sont susceptibles de recevoir, les archives des écrivains contemporains représentent, par leur originalité, leur diversité et leur amplitude, un défi pour celles qui acceptent d'assumer leur conservation, leur préservation et de faciliter leur consultation. Du point de vue de la recherche, au sein des humanités numériques en particulier, ces archives constituent une importante réserve de données à condition, bien sûr, que celles-ci soient exploitables. Cet article propose d'expliquer comment, depuis 2003, nous nous sommes engagés dans la constitution d'une telle réserve à travers le projet de recherches et de valorisation du fonds « Alcides Giraldi ». Ce fonds, actuellement hébergé à Lille, comporte de nombreux inédits susceptibles de documenter la recherche internationale sur la circulation des idées entre l'Europe et les Amériques durant les principaux conflits du XX<sup>ème</sup> siècle.

## Introduction

Le patrimoine écrit est la raison d'être des bibliothèques et des archives, publiques et privées : en elles se conservent les aventures de la pensée, les garants de la parole donnée, le trésor de la langue et la mémoire des nations.

Pierre-Marc DE BIASI, « Le patrimoine écrit »<sup>3</sup> (2007).

Parmi les patrimoines écrits que les institutions publiques ou privées sont susceptibles de recevoir, les archives des écrivains contemporains représentent, par leur originalité, leur diversité et leur amplitude, un défi pour celles qui acceptent d'assumer leur conservation, leur préservation et de faciliter leur consultation. Du point de vue de la recherche, au sein des humanités numériques en particulier, ces archives constituent une importante réserve de données à condition, bien sûr, que celles-ci soient exploitables. Cet article propose d'expliquer comment, depuis 2003, nous sommes engagés dans la constitution d'une telle réserve à travers le projet de recherches et de valorisation du fonds « Alcides Giraldi » qui comporte des écrits inédits susceptibles de documenter la recherche internationale sur la circulation des idées entre l'Europe et la région du Río de la Plata durant le XX<sup>ème</sup> siècle (en particulier durant la Guerre Civile espagnole, la Seconde Guerre Mondiale et les dictatures du Cône Sud.). Notre travail, destiné à mettre ces documents à la disposition de tous, a démarré par le classement, le conditionnement et la préservation de ces fonds. Actuellement, nous préparons l'exploitation et la diffusion numérique de ces sources par la constitution d'une base de données et de métadonnées interopérables. Les pages qui suivent proposent, à travers l'état des lieux de ce programme, d'interroger certaines de nos options passées et à venir, à l'aune des réflexions actuellement menées par les chercheurs en « Humanités Numériques ». Après avoir présenté les fonds d'archives au cœur de notre programme, nous expliquerons quels ont été nos choix méthodologiques pour traiter ces manuscrits et promouvoir les recherches sur ceux-ci, nous décrirons dans le même temps les limites de ces options liminaires et présenterons quelques unes des pistes envisagées pour la suite, en vue de l'extraction et de la diffusion de ces connaissances.

---

<sup>3</sup> Pierre-Marc DE BIASI, « Le patrimoine écrit », url : <http://www.item.ens.fr/index.php?id=13597>.

## 1. Un projet patrimonial

### 1.1 La sauvegarde et la préservation des manuscrits

Les archives concernées par ce programme appartiennent à deux fonds uruguayens : le fonds « Alcides Giraldi » (il s'agit des documents et manuscrits des écrivains José Mora Guarnido et Carlos Denis Molina) et le fonds « Carlos Liscano ». Tous deux sont actuellement en dépôt au Service Commun de la documentation de Lille 3 où ils ont été transférés entre 2003 et 2009 grâce au soutien du Centre d'Études et de recherches sur les Civilisations, les Langues et Littératures Étrangères (CECILLE EA 4074, Lille 3). La signature d'une convention (renouvelée tous les cinq ans), entre les ayants-droits (héritiers « Alcides Giraldi » et l'écrivain Carlos Liscano) et l'Université de Lille 3 autorise l'exploitation scientifique de ces documents, la publication des résultats de ces recherches ainsi que les projets éditoriaux sur ces manuscrits. Ces fonds, dont une présentation détaillée est consultable en ligne sur le site internet dédié au projet<sup>4</sup>, comportent des photos, des enregistrements sonores, des lettres manuscrites, des articles de presse, des cours tapuscrits et des brouillons d'œuvres littéraires inédites consignés dans les carnets manuscrits ou des notes éparses. L'ensemble de ces documents témoigne, sous différentes formes, de l'effet des mobilités intellectuelles entre l'Europe et les Amériques durant les principaux conflits internationaux et civils du siècle dernier<sup>5</sup>.

L'une des premières vocations de notre travail sur ces archives, a été, et demeure, patrimoniale : en effet, durant près de quinze ans, les documents ont dû être cachés dans des conditions précaires pour être protégés des forces armées qui ont occupé le pouvoir en Uruguay durant la dictature (1972-1985); leur arrivée à Lille a donc permis de les sauver d'une inéluctable détérioration. Après une quarantaine de quelques mois,

---

<sup>4</sup> Voir le site <http://manuscritsentredeux.recherche.univ-lille3.fr/> et voir également quelques articles qui présentent ces fonds : Fatiha Idmhand, « Le fonds José Mora Guarnido » et Cécile Braillon-Chantraine, « Fonds Carlos Denis Molina », in *Catalogue des collections patrimoniales*, Université Lille 3, 2009, p. 94-98 et p.99-104.

<sup>5</sup> Comme nous l'avons montré dans différents travaux de recherches et articles publiés sur ces archives comme par exemple dans : Fatiha Idmhand, *Échanges en exil, exils en échanges : la correspondance de José Mora Guarnido* in Ada Savin (ed.) *Migrations and exile: Charting New Literary and Artistic Territories*, Cambridge Scholars Publishing, 2013 (ISBN 1-4438-4402-0), p. 47-56 ; ou dans cet article collectif Eleonora Basso / Carlos Demasi / Norah Giraldi Dei Cas / Fatiha Idmhand « Trayectoria de José Mora Guarnido. Espejo de un intelectual entre España y América (1923-1939) » in Mora, Carmen de / García Morales, Alfonso (eds.), *Viajeros, diplomáticos y exiliados. Escritores hispanoamericanos en España (1914-1939)*, Tomo I y II, P.I.E. Peter Lang, Bern, Berlin, Bruxelles, Frankfurt am Main, New York, Oxford, Wien, 2012 (ISBN 978-90-5201-814-0) ; Tomo II, p.517-539.

mise à profit pour trier les documents, éliminer les conditionnements initiaux qui représentaient un danger pour leur conservation<sup>6</sup>, nous avons pu préparer l'organisation intellectuelle du fonds ainsi que le plan de classement en vue des études à venir. Cette première approche a été codicologique et génétique : elle nous a permis de reconstituer les œuvres dispersées ainsi que la chronologie des documents grâce à l'analyse des indices matériels, du papier, de sa nature, de la couleur de l'encre et/ou de la forme de l'écriture: des traces de ce travail sont présentes dans les rubriques des notices bibliographiques que nous concevons comme nous aurons l'occasion de l'expliquer dans les pages qui suivent.

A chacune de ces étapes, la collaboration entre le chercheur et l'archiviste a été permanente et a permis d'entrevoir la plupart des options qui ont été retenues ensuite pour la préservation des documents et pour leur numérisation. En effet, en raison de la fragilité des documents, nous avons très vite envisagé leur numérisation afin de limiter les manipulations, tout en permettant l'exploitation scientifique de ceux-ci. Dès 2007, nous avons donc procédé, à une série de numérisation « de sauvegarde » des pièces les plus fragiles (la presse et la correspondance notamment). La réalisation de ce travail a été confiée à un prestataire extérieur (Atelier National de Reproduction des Thèses, Lille<sup>7</sup>) afin de préserver l'intégrité des pièces et la numérisation a été effectuée en haute résolution<sup>8</sup> en vue de la diffusion en ligne. Chaque document a été nommé avec le code alphanumérique identifiant de l'institution, le code du fonds et le code du document et le résultat de ce premier effort est aujourd'hui consultable sur le site <http://manuscritsentredeux.recherche.univ-lille3.fr>.

## 1.2 La communication autour des fonds

L'ouverture d'un site web destiné au signalement international et à la consultation de ces fonds nous a toujours semblé nécessaire car nous avions à l'esprit l'intérêt, pour des chercheurs d'autres pays du monde, d'avoir accès à ces documents. Faute de financement régulier, il ne nous a pas encore été possible de proposer un centre de ressources numériques performant ; nous nous sommes néanmoins efforcés de proposer une première ébauche destinée à « communiquer » sur le projet et sur ces fonds par la création, en 2009-2010, du site <http://manuscritsentredeux.recherche.univ-lille3.fr>.

---

<sup>6</sup> Nous avons supprimé les reliures en métal, les pochettes humides et cordelettes, acheté du matériel spécifique pour l'archivage selon les recommandations de l'état de l'art dans le domaine et grâce à l'aide de Cécile Martini, archiviste-paléographe. Voir par exemple ce document publié par l'ENSSIB : *La conservation des documents papier. Point sur l'évolution des techniques et des stratégies*, in <http://bbf.enssib.fr/consulter/bbf-2006-04-0011-002.pdf>, 2006.

<sup>7</sup> L'ANRT dispose d'un service dédié à la numérisation d'ouvrages anciens et précieux. Leur site : <http://www.diffusiontheses.fr/>

<sup>8</sup> 600 dpi couleur, TIFF, JPEG, et une conversion en PDF pour pouvoir travailler sur ces manuscrits chez nous.

Basé sur le CMS Spip, un système de publication facile d'emploi, libre d'accès, rapide à mettre en place, le site offre, en dépit d'une « ergonomie » moyenne<sup>9</sup>, une vue globale de l'arborescence grâce à des menus organisés en rubriques en administration, une navigation élémentaire avec une rubrique *Présentation du projet de recherche*, une partie éditoriale et, surtout, les *Archives en ligne* consultables grâce à un outil de visualisation « Flippingbook » qui permet de feuilleter les archives numériques et de zoomer sur la page<sup>10</sup>. Trois ans après son ouverture, malgré ses limites, cette première version du site a tout de même atteint deux principaux objectifs : elle nous a permis de révéler l'existence de ces fonds, de faire connaître notre projet et de susciter l'intérêt de chercheurs d'Espagne et des Amériques grâce à la consultation de documents en ligne, ensuite, elle a pu servir de faire-valoir auprès d'institutions susceptibles de nous aider à développer ce programme (telles que la Maison des Sciences de l'Homme, MESHS de Lille ou l'Agence Nationale de la Recherche) afin d'obtenir des financements et pérenniser le projet. Néanmoins, comme nous le montrerons plus loin dans ce travail, il est aujourd'hui nécessaire de revoir ce modèle, afin d'y intégrer notamment des fonctionnalités plus élaborées pour une valorisation optimale des corpus.

### 1.3 La génétique : une science pour les manuscrits

Pour l'analyse critique et littéraire de ces manuscrits, nous avons adopté la méthodologie « génétique » telle qu'elle est définie et développée par les chercheurs de l'Institut des Textes et Manuscrits Modernes (ITEM, CNRS/ENS, UMR-8132). Née en France dans les années 1970, cette science a renouvelé la connaissance des textes en intégrant à la compréhension de ceux-ci les brouillons de l'œuvre. En effet, alors que la nouvelle critique avait en partie délaissé la figure de l'auteur, voire « tué » celle-ci (R.Barthes)<sup>11</sup> dans les années 1960, la « critique génétique » en a proposé une nouvelle lecture par l'analyse du processus de l'écriture (L. Hay<sup>12</sup>). L'approche critique et génétique sonde donc la mémoire de l'acte de création en collectant les « traces » et les indices matériels inscrits dans les brouillons pour comprendre la conception, la préparation et la rédaction de l'œuvre<sup>13</sup>. « Empirique

---

<sup>9</sup> Comme le signale Erwan Tanguy sur son blog, « Spip a, du coup, une réputation de sites moches mais fonctionnels » à lire sur : <http://blog.voyelle.me/cms/wordpress-vs-spip>.

Voir également la page : <http://www.idd.fr/solutions/spip-le-cms-a-tout-faire.html> et le site officiel [http://www.spip.net/fr\\_rubrique91.html](http://www.spip.net/fr_rubrique91.html).

<sup>10</sup> Flipping Book PDF Publisher Basic, <http://flippingbook.com/>.

<sup>11</sup> Roland Barthes, *Le Bruissement de la langue*, Paris, Seuil, coll. Points Essais, 1993, chap. « La mort de l'auteur », p. 63-69.

<sup>12</sup> Voir l'article de Louis Hay, « Qu'est-ce que la critique génétique ? », publié en ligne sur le site de l'ITEM : <http://www.item.ens.fr/index.php?id=44566>

<sup>13</sup> Pierre-Marc de Biasi, *La génétique des textes*, Paris, Nathan PUF, 2003.

» (De Biasi) et phénoménologique, cette méthode permet donc de dégager des significations à partir des métamorphoses du manuscrit et des étapes de l'élaboration de l'écriture. C'est ainsi que dans le cas des manuscrits de José Mora Guarnido, par exemple, l'étude de ses brouillons et l'analyse de ses multiples campagnes de corrections et ratures permettent de comprendre comment, au fil des ans, une mémoire de l'espace et du pays perdu s'installe dans les textes ; chez Carlos Denis Molina, cette approche permet, en remontant le fil d'une transformation générique grâce aux cahiers manuscrits (passage du roman au théâtre par exemple), de déceler les influences du théâtre avant-gardiste.

## **2. Le traitement de ces archives : enjeux du traitement des fonds littéraires contemporains**

### **2.1 La description des manuscrits**

Comme souvent lorsqu'une bibliothèque reçoit le fonds d'un écrivain contemporain, ses difficultés résident en ce qu'elle doit le traiter à la fois comme fonds d'archives et comme fonds documentaire. La réflexion sur les normes de description de ces archives a donc représenté un enjeu à la fois pour le signalement mondial des fonds mais également pour la production de nos métadonnées. En effet, non seulement « il existe une confusion terminologique entre le monde des archives et celui des bibliothèques, sur la notion même de manuscrits »<sup>14</sup> mais l'hétérogénéité du fonds de l'écrivain contemporain rend sa description complexe. Dans l'élaboration de la structure de la base de données, cela suppose par conséquent que la description du fonds comporte des éléments qui relèvent de domaines distincts de la bibliothéconomie. Si, par ailleurs, on veut respecter les normes et principes traditionnels des catalogues des bibliothèques, cela veut dire qu'en tant que fonds d'archives, le fonds doit être décrit depuis l'archivistique et qu'en tant que fonds documentaire, il faut en décrire chacune des unités, lesquelles composent le corpus d'archives. Ces croisements ont exigé de nous une réflexion sur une forme « personnalisée »<sup>15</sup> de la description et du catalogue.

---

<sup>14</sup> Agnès Vatican : « dans les archives, les documents manuscrits sont prédominants, mais ne relèvent pas tous de la définition qu'on donne dans les bibliothèques, plus proche d'œuvre de l'esprit » in « Les manuscrits littéraires dans les archives en France », article publié in Martine Sagaert, *Manuscrits littéraires du XX<sup>e</sup> siècle*, Presse Universitaire de Bordeaux, 2005, p. 67 ; cité par Coline Monkemeier, *Traitement, valorisation et description du fonds José Mora Guarnido*, Rapport de stage Licence Pro « Chargé de valorisation des ressources documentaires », Mission effectuée du 15 avril au 10 septembre 2010, au Service Commun de la Documentation de l'Université Lille 3.

<sup>15</sup> Voir le chapitre « Signaler et faire connaître des documents patrimoniaux », in Raphaëlle Mouren, *Manuel du patrimoine en bibliothèque*, Paris, Éditions du Cercle de la Librairie, 2007, cité par Coline Monkemeier, *ibid.*



La publication, en 2010, des recommandations pour la *Description des manuscrits et fonds d'archives modernes et contemporains en bibliothèque* est tombée à point nommé puisqu'elle est advenue au milieu de notre travail sur le plan de classement et qu'elle a coïncidé avec le recrutement de deux stagiaires chargés d'élaborer la base de données qui allait servir à décrire nos fonds. L'initiative de Direction du livre et de la lecture, au sein de la Commission de normalisation Modélisation, production et accès aux documents, est née d'un constat que nous avions nous-mêmes fait : la norme pour la description des archives qui était utilisée par de nombreuses bibliothèques dont les catalogues étaient informatisés, l'ISAD(G)<sup>16</sup>, ne convenaient pas au traitement des fonds littéraires et aux spécificités de ce type de documents<sup>17</sup>. Le groupe de travail réuni sous la direction de l'AFNOR et de la Direction du livre et de la lecture, a donc préparé, en 2010, une série de recommandations, le plan DeMarch<sup>18</sup>, destinées à « harmoniser les pratiques dans le contexte de l'informatisation des catalogues tout en prenant en compte les spécificités des fonds conservés dans les bibliothèques »<sup>19</sup>, notamment les fonds littéraires. Ce document, qui reprenait à son tour d'autres recommandations internationales, nous a donc permis de travailler dans le respect de normes qui permettraient la circulation de nos informations. C'est pourquoi nous avons intégré ces critères à l'élaboration de notre base de données sous Notix.

---

<sup>16</sup> Traduction française : <http://www.ica.org/10225/normes/isadg-norme-gnrale-et-internationale-de-description-archivistique-deuxime-dition.html>

<sup>17</sup> Voir *Les instruments de recherche dans les archives* ; Ministère de la Culture. Direction des Archives de France, Paris : documentation française, 1999, p. 91-92. Source citée par Aurélie Delamarre, dans son mémoire d'études pour le Diplôme de conservateur de bibliothèque, *Traitement et catalogage des manuscrits contemporains : défis et enjeux. L'exemple de la bibliothèque municipale d'Albi, janvier 2004*, en ligne sur : <http://www.enssib.fr/bibliotheque-numerique/documents/1033-traitement-et-catalogage-des-manuscrits-contemporains.pdf>, p.55.

<sup>18</sup> « Le groupe de travail s'est inspiré de la norme Describing Archives : a content standard (DACS) qu'il a traduit et adaptée aux usages des établissements français. Il s'est également référé ponctuellement à d'autres normes existantes pour la description des ressources archivistiques (Règles pour la description des documents d'archives (RDDA) par ex. ) et, le cas échéant, aux normes existantes pour des types de documents précis (archives de films par ex.) ». Voir Association française de normalisation. Commission de normalisation Modélisation, production et accès aux documents, *Description des manuscrits et fonds d'archives modernes et contemporains en bibliothèque : DeMArch* [en ligne], 2010, <http://www.bivi.fonctions-documentaires.afnor.org/livres-blancs/recommandation-de-description-des-manuscrits-et-fonds-d-archives>. Voir également, sur le site de la BnF, le projet en ligne : [http://www.bnf.fr/fr/professionnels/normes\\_catalogage/a.ead\\_demarch.html](http://www.bnf.fr/fr/professionnels/normes_catalogage/a.ead_demarch.html).

<sup>19</sup> <http://bonnespratiques-ead.net/guide/intro/autres-normes>

## 2.2 Notix : un programme pour la description des fonds

Notix est un système de gestion de bases de données documentaire conçu pour constituer (et maintenir) des catalogues bibliographiques dans un environnement réparti entre plusieurs contributeurs<sup>20</sup>. Il fonctionne en mode web et permet la saisie, la consultation de notices bibliographiques ainsi que le traitement de documents XML natifs grâce à un système de gestion de données XML couplé aux moteurs de recherche Lucene<sup>21</sup> et XML eXist<sup>22</sup>. Une fois paramétré, l'outil permet de disposer de notices et de métadonnées structurées. Par ailleurs, il offre à l'utilisateur des fonctionnalités d'administration des bases, de recherches et de consultation simples et efficaces.

Le paramétrage de cet outil étant entièrement personnalisable, nous nous sommes proposés de l'adapter afin d'associer les dimensions pragmatiques et prescriptives de la bibliothéconomie (description du document, cotation...) à celles, plus circonstanciées, de nos études critiques et génétiques. Pour établir le modèle de nos notices, définir les principaux champs de celles-ci, et le plan de classement incarnant cette arborescence, nous nous sommes inspirés des bases de données d'autres institutions nationales ou européennes hébergeant des manuscrits contemporains telles que la BnF, l'IMEC (Institut des Textes et Manuscrits Modernes), la Bibliothèque Jacques Doucet ou les Archives littéraires suisses (Bibliothèque nationale suisse) et, surtout, des recommandations du plan DeMArch (2010). L'image qui suit, illustre le résultat de ce travail et le modèle auquel nous sommes parvenus :

---

<sup>20</sup> Le choix de cette base de données documentaire est notamment lié à l'intégration, au sein de l'équipe, d'André Davignon, qui avait participé, entre 2005 et 2006, au sein du Ministère de l'Écologie, de l'Énergie, du Développement Durable et de l'Aménagement du Territoire, à l'élaboration de ce programme.

<sup>21</sup> L'application s'appuie sur le SGBD XML eXist, le moteur de recherche Lucene et la plateforme de publication XML Cocoon. Voir <http://lucene.apache.org> ou <http://exist.sourceforge.net/>

<sup>22</sup> Les sources de Notix (<http://notix.org>) ont été déposées sur Adullact, la plateforme collaborative des administrations françaises pour le développement de logiciels libres(<https://adullact.net/projects/notix/>).

Production de données en humanités numériques : l'exemple du « Fonds Alcides Giraldi »

## Guarnido

JMG AA 1.1 [CA] Don Pedro C1

Localisation	JMG AA 1.1 [CA] Don Pedro C1
Recueil	Cuentos Americanos
Titre propre	Don Pedro el melancólico
Auteur	José Mora Guarnido
Date de début	1961
Dates extrêmes	1961
DESCRIPTION	
Support	Feuillets
Type	Tapuscrit
Volume	5 feuillets numérotés de 2 à 24
Format	318 mm x 224 mm
Etat général	Bon
Reliure	Feuillets perforés
Etat des feuillets	Bon
Etat des feuillets - Note	Le premier feuillet comporte le titre, de façon manuscrite au stylo noir, en haut à droite (écriture tremblante).
Couleur de l'encre	Noire
Corrections	Rares
Couleur des corrections	<ul style="list-style-type: none"> <li>• Encre noire</li> <li>• Encre bleue</li> </ul>
Notes	Le dernier feuillet comporte la date tapuscrite : "Fevrero 1961".
Numérisation	Non
DESCRIPTION ANALYTIQUE	
Thème	exil
Genre	Récit Bref
Référent spatial	Montevideo
Hiérarchie géographique	<ul style="list-style-type: none"> <li>• Amérique</li> <li>• Uruguay</li> </ul>
Publication	Non
RENOIS DANS LE FONDS	
Renvoi par dossier génétique	<ul style="list-style-type: none"> <li>• JMG AA 1.1 [CA] Don Pedro C2</li> <li>• JMG A 9.1 [CA] Don Pedro</li> </ul>
Analyse	Projet de José Mora Guarnido intitulé Cuentos americanos et qui rassemble trois récits brefs, tous tapuscrits, dont l'action se passe à Montevideo à une époque contemporaine à l'auteur. Ces récits ont un point commun, ils évoquent l'immigration espagnole en Amérique latine et surtout en Uruguay.
Mentions légales	Fonds ALCIDES - GIRALDI Archives José Mora Guarnido Montevideo / Lille

FIG. 1 - Notice bibliographique générée par Notix.

Fonds José Mora Guarnido, Manuscrit de « Don Pedro el melancólico ».

La notice montre, en plus des rubriques traditionnelles « localisation, titre, recueil, dates... », une série de champs destinés à faciliter (ou anticiper) le travail (à venir) du chercheur lors de sa consultation des fonds. Dans la mesure où Notix permet de croiser les recherches, nous avons ajouté une série de champs qui permettent de situer l'œuvre au sein de son « dossier génétique » et reconstituer le parcours créatif. L'idée des rubriques « Renvoi par dossier génétique » et « Analyse », est d'offrir à l'utilisateur quelques informations sur la genèse du projet, et de lui permettre d'accéder, facilement, à l'ensemble des copies, manuscrits ou documents liés à l'œuvre consultée.

Les recherches croisées par année, nom, thème, type de manuscrit (carnet, feuillets...) ou état des manuscrits permettent, en outre, de constituer son propre corpus. Ainsi par exemple, ce paramétrage « à la carte » permet, face à des manuscrits complexes comme les carnets de notes de José Mora Guarnido, de tracer les multiples relations d'un document avec d'autres. En effet, les carnets de José Mora Guarnido comportent souvent plusieurs projets littéraires car l'écrivain avait pour habitude de commencer la rédaction d'une œuvre dans un sens, de poursuivre celle-ci dans un autre et d'entamer parfois la rédaction d'une autre au milieu des pages avant de continuer, si besoin était, la rédaction dans un autre carnet.<sup>23</sup> Les rubriques « analyse » et surtout « renvoi par dossier génétique » permettent alors de situer le projet au milieu des autres brouillons d'œuvres et, grâce à rubrique « renvoi par dossier génétique », d'associer le document consulté aux autres documents du fonds liés à celui-ci (autres copies, manuscrits, carnets, articles de presse, correspondances, photos ou enregistrements...). La notice générée (sous XML, PDF ou ODT) offre donc à l'utilisateur une information complète sur les spécificités et la thématique du document consulté : depuis sa « localisation » dans le fonds, son « Thème », son « Référent spatial », jusqu'à son « état » et des informations sur l'encre utilisée, une liste qui n'est évidemment pas exhaustive, mais qui témoigne par ailleurs de l'état des recherches actuelles sur les fonds d'archives (Fonds José Mora Guarnido, pour le cas de la notice ci-dessus) et qui reste ouverte à d'autres propositions<sup>24</sup>.

Si la rédaction de ces informations présuppose qu'un chercheur ait repéré ces corpus et préparé, en amont, les dossiers de genèse, les associations en question sont matérialisées, dans la base, par des relations RDF et des hyperliens qui pointent vers les notices associées. Les métadonnées saisies matérialisent donc à la fois la description archivistique du corpus et apporte des éléments d'analyses sur les relations et liens (logiques, temporels, etc.) qu'entretiennent les différents documents du fonds.

La réflexion sur l'élaboration de ces champs est le fruit du travail de l'ensemble des membres de l'équipe (archiviste, informaticien, webmestre, chercheurs et étu-

---

<sup>23</sup> On trouve parfois des mentions manuscrites telles que celle-ci : « *Ver I Creo que hay copia de esto* » traduction : « *Voir I Je pense qu'il y a une copie de tout cela* ».

<sup>24</sup> Pour le cas de José Mora Guarnido, nous avons tenu compte d'un projet de plan de classement que l'écrivain avait laissé parmi ses archives.

dians). Cette entreprise collective a permis de consigner, dans les rubriques, les métadonnées nécessaires pour structurer l'information et décrire les différents objets du fonds : métadonnées descriptives, administratives, de gestion (gestion des droits par exemple), des métadonnées de préservation et des données techniques décrivant les caractéristiques physiques de ces sources. Surtout, ces métadonnées respectent les normes de présentation, de mise en forme, de mise en page, de ponctuation et de description préconisées par les principales infrastructures nationales et internationales : il s'agit pour nous d'assurer leur interopérabilité et leur inscription dans les catalogues en ligne de nombreuses bibliothèques grâce aux « moissons » réalisées par les principaux portails scientifiques francophones (ISIDORE, GALLICA, moteurs de recherche..) et européens (EUROPEANA). En outre, la description de chaque unité documentaire avec des métadonnées Dublin Core permet l'inscription de ces données dans les entrepôts OAI-PMH.

### 2.3 Limites de Notix

Pour les fonds d'archives qui nous concerne, Notix est, pour l'instant, un outil pertinent, qui facilite l'association de deux dimensions de notre projet : la bibliothéconomie, d'une part, l'étude critique et génétique de ces manuscrits, de l'autre. Néanmoins, ce programme pose deux problèmes : l'un concerne les compétences « sérieuses »<sup>25</sup> qu'il exige pour son « installation et son paramétrage »<sup>26</sup>, ce qui peut poser des difficultés lorsque les équipes ou les moyens sont réduits ou/et régulièrement compromis comme c'est le cas pour ce projet ; l'autre concerne l'adaptation de ce système à l'EAD, la norme utilisée en France pour encoder les manuscrits (anciens et/ou orientaux) et pour faciliter la diffusion en ligne de ces corpus. En effet, même si Notix n'est pas incompatible avec cette norme XML, elle-même entièrement compatible avec la norme ISAD(G)<sup>27</sup> dans sa version actuelle (2002), celle-ci s'impose en France pour devenir le « format de catalogage courant pour la description des fonds d'archives et des manuscrits, aussi bien dans les services d'archives que dans les bibliothèques »<sup>28</sup>. Maintenant utilisé pour décrire les manuscrits conservés à la BnF et par le réseau Calames pour les bibliothèques universitaires, nous nous interrogeons en ce

---

<sup>25</sup>Voir : [http://www.bibliofrance.org/index.php?option=com\\_content&view=article&id=195:notix-un-nouveau-logiciel-libre-pour-les-centres-de-documentation-et-bibliotheques&catid=52:ressources&Itemid=53](http://www.bibliofrance.org/index.php?option=com_content&view=article&id=195:notix-un-nouveau-logiciel-libre-pour-les-centres-de-documentation-et-bibliotheques&catid=52:ressources&Itemid=53)

<sup>26</sup>*Ibid.*

<sup>27</sup> « Chaque élément de description archivistique défini par l'ISAD(G) est facilement exprimé en EAD, soit par un élément spécifique de l'EAD, soit par une combinaison d'éléments. En outre, grâce à la structure arborescente du langage XML et à la gestion de l'héritage qu'il permet, l'EAD respecte parfaitement les quatre règles de la description à plusieurs niveaux définies par l'ISAD(G) », in <http://bonnespratiques-ead.net/guide/intro/autres-normes>

<sup>28</sup>Url : [http://www.bnf.fr/fr/professionnels/formats\\_catalogage/a.f\\_ead.html#SHDC\\_\\_Attribute\\_BlocArticle3BnF](http://www.bnf.fr/fr/professionnels/formats_catalogage/a.f_ead.html#SHDC__Attribute_BlocArticle3BnF).

moment sur la façon d'associer Notix au format EAD et sur les métadonnées descriptives, administratives et structurelles (type standard METS de la Library Of Congress) qu'il faudra intégrer à la description de nos unités documentaires. Globalement, cela suppose d'intégrer à l'équipe de nouvelles compétences pour procéder aux conversions de ces données ou de rebrousser chemin?<sup>29</sup>

### 3. Enjeux de la valorisation des sources

#### 3.1 Diffuser des connaissances pour promouvoir la recherche en Humanités Numériques

La feuille de route des Très Grandes Infrastructures de Recherche a inscrit parmi ses priorités la numérisation et l'indexation des sources pour contribuer au développement des savoirs sur les cultures et les civilisations: notre programme sur ces fonds d'archives uruguayens concourt donc à un large mouvement qui veut faciliter l'accès à des sources méconnues et diffuser, comme nous l'avons dit plus haut, des documents qui éclairent les connaissances sur la circulation de l'information, de part et d'autre de l'Atlantique, durant les périodes de conflits du siècle dernier. Face au « tournant numérique » actuel, l'idée de notre programme est aussi de participer aux réflexions actuellement en cours au sein des sciences humaines et sociales, et de la littérature, sur la façon dont ces mutations affectent nos objets d'études. En effet, l'impact de ce tournant dans notre domaine, celui de la genèse de l'écrit, de l'analyse des processus à l'œuvre dans les brouillons, de l'étude des manuscrits et archives des écrivains, est tout à fait remarquable. En cette fin de l'ère Gutenberg<sup>30</sup>, le papier n'est plus la mémoire exclusive des « traces » sur lesquels nous avons l'habitude de fonder nos analyses et théories : la nature de l'archive est en train d'évoluer avec les pratiques de la société et des auteurs, toujours plus mobiles et connectés. L'écrivain contemporain, ses écrits, la nature des traces qu'il laisse et même l'objet « livre » qu'il publie, ont changé : on constate ainsi des différences importantes entre des auteurs comme Patricio Pron (Argentine, 1976- ) qui, depuis qu'il a dû quitter son pays, l'Argentine, lors de la crise de 2002, a décidé de brûler ses manuscrits pour « voyager léger »<sup>31</sup>; ceux qui investissent l'internet pour y concevoir leurs textes, les diffuser, promouvoir et même y composer leurs propres « archives numériques » comme l'écrivain espagnol Robert Juan-Cantavella<sup>32</sup>; ou ceux qui, comme le belge Jean-Philippe Toussaint, documentent la

---

<sup>29</sup> La réflexion est en cours sur ce point tandis qu'un peu moins de deux mille notices ont été saisies.

<sup>30</sup> J'emprunte la notion à Enrique Vila Matas, *Dublinesca*, Barcelona, Seix Barral, 2010.

<sup>31</sup> Non sans en avoir conservé une copie scannée comme il l'expliquait dans une conférence donnée à Lille : <http://www.univ-lille3.fr/actualites/?actu=7010>

<sup>32</sup> Journaliste de formation et très actif sur le net avec des blogs (<http://asesinocosmico.blogspot.fr/>) et des publications en ligne, l'écrivain a livré au public les

genèse de ses textes en déposant sur sa propre page internet, les témoins génétiques de ses romans<sup>33</sup>. Sans parler des écrivains dont les productions sont nativement numériques! L'un des auteurs issus du corpus au cœur de notre projet, Carlos Liscano (Uruguay, 1949- ), dispose d'archives qui vont des manuscrits de prison<sup>34</sup> à la copie dactylographiée en passant par la disquette 3,5 pouces, le DVD, la clé USB, le disque dur externe, l'édition sur Facebook et le compte Dropbox, un matériau qui témoigne en lui-même de la grande *conversion numérique*<sup>35</sup> que nous vivons depuis quelques années.

Comme nous l'avons montré, cet « univers archival hybride »<sup>36</sup> qui voit évoluer, simultanément, les corpus et matériaux traditionnellement au cœur des « Humanités », ainsi que les outils qui ont servi à les analyser jusqu'à présent, ouvre de nouvelles possibilités grâce à la diversification des pratiques, au travail *transdisciplinaire*<sup>37</sup> et à la publication en ligne. En effet, si nous avons beaucoup insisté dans les pages qui précèdent, sur les différentes normes et recommandations dont nous avons tenu compte pour la rédaction et la préparation de nos métadonnées (afin de faciliter leur circulation et leur référencement), notre but n'est pas seulement de transférer des sources d'un format à l'autre mais bien de faciliter la recherche scientifique par l'exploitation numérique de celles-ci. Alors que la comparaison des différentes versions d'un même texte est facilitée par l'édition en ligne<sup>38</sup>, l'enrichissement de celle-ci par des balises significatives permet de concrétiser le projet de recherche.

### 3.2 La TEI pour décrire et publier en ligne

Depuis les années 1990, le projet Text Encoding Initiative (TEI)<sup>39</sup> s'est proposé de construire le schéma facilitant l'enrichissement des éditions en ligne. Très orientée

---

documents de genèse (photos, notes, articles de presse, manuscrits...) de son livre *El dorado* sur un site accessible à tous <http://punkjournalism.net/>.

<sup>33</sup> L'écrivain a mis en ligne les manuscrits de deux de ses œuvres en les classant selon leurs états génétiques : <http://www.jptoussaint.com/france.html#cahiers-d-archives>

<sup>34</sup> Voir l'édition en ligne de ces manuscrits sur <http://manuscritsentreux.recherche.univ-lille3.fr/archives-manuscrits-en-ligne/carlos-liscano/fonds-d-archives-carlos-liscano/article/consulter-les-inedits>

<sup>35</sup> Pour plagier le titre de l'ouvrage du philosophe Milad Douehi, *La grande conversion numérique*, Paris, Seuil, 2008.

<sup>36</sup> *Ibid.*, p.212.

<sup>37</sup> <http://tcp.hypotheses.org/318>

<sup>38</sup> « Digital environments provide the ability to pull together many versions of a single work, tracking its development, noting its variants, and presenting the whole comparative array of witnesses », in Anne Burdick, Johanna Drucker, Peter Lunenfeld, Todd Presner, Jeffrey Schnapp, *Digital Humanities*, The MIT Press, Massachusetts, 2012, p.35.

<sup>39</sup> Url : <http://www.tei-c.org/index.xml>

vers la publication Internet, cette forme de DTD offre une importante bibliothèque de balises constituant un cadre méthodologique<sup>40</sup> relativement personnalisable qui permet de décrire les spécificités concrètes et matérielles des manuscrits ainsi que les phénomènes littéraires et génétiques<sup>41</sup> (citations, métaphores, suppressions, ratures...). La vocation littéraire de la TEI, dont l'objectif premier n'est pas l'analyse textométrique stricto sensu, mais bien l'approche critique, nous a conduit à envisager ce système de balisage. Même si les transcriptions de manuscrits sur lesquels nous travaillons actuellement n'ont pas encore fait l'objet d'édition en ligne ou d'encodage, les formations suivies et les résultats scientifiques obtenus par les projets qui utilisent la TEI<sup>42</sup> dans leur approche, nous confirment que les possibilités offertes par ce langage répondent aux principales ambitions d'un programme en Humanités Numériques : le format est un standard international et son schéma d'encodage en XML permet la circulation de l'information or, comme le rappelle Milad Douehi, l'interopérabilité et la compatibilité sont « les caractéristiques de base de tout cadre concevable pour l'archivage numérique »<sup>43</sup>.

L'encodage par la TEI permet de documenter l'édition électronique de différentes versions d'un même tapuscrit, en commentant ces campagnes de corrections et de réécritures. Face à ces centaines d'inédits, nous espérons pouvoir fixer des textes et en proposer une édition qui n'effacerait pas l'histoire de la genèse. On peut même imaginer, pour certains types de documents, solliciter la contribution de l'internaute dans une démarche participative.<sup>44</sup> Enfin, l'autre élément qui nous a poussés à opter pour la TEI est qu'elle permet également d'encoder les archives sonores : nous disposons d'une série d'enregistrements audio de programmes radios de Carlos Denis Molina datant des années 1950 qui pourront ainsi être séquencées et synchronisées avec du contenu scientifique.

---

<sup>40</sup> Url : <http://www.revue-texto.net/Corpus/Manufacture/standards/d1e284.html>

<sup>41</sup> Cours de Lou Burnard (Oxford Université) sur la façon dont la TEI peut être utilisée pour encoder des textes contemporains, de l'essai à la fiction, dans l'optique d'une édition multi-supports (papier et électronique) : <http://leo.hypotheses.org/2630>

<sup>42</sup> Comme les projets en cours de développement au sein du Labex Obvil (<http://www.paris-sorbonne.fr/labex-obvil>) ou le projet des Bibliothèques Virtuelles Humanistes de Tours (<http://www.bvh.univ-tours.fr/manuscrits.asp>) ou les projets d'éditions critique et génétique des manuscrits de Marcel Proust ([http://research.cch.kcl.ac.uk/proust\\_prototype/about.html](http://research.cch.kcl.ac.uk/proust_prototype/about.html)) ou de Juan Goytisolo (<http://dh13.sagw.ch/fr/dh13/poster/poster/vauthier-camus-sager.html>) et <http://goytisolo.unibe.ch/>)

<sup>43</sup> Milad Douehi, *op. cit.*, p.212.

<sup>44</sup> Pour la navigation, l'exploration, l'extraction et, surtout, la visualisation des données, nous nous inspirerons de la démarche des projets Proust et Goytisolo qui ont inclus dans leurs prototypes, des outils de comparaison par le biais de SVG ou Joomla, entre la vue numérisée et sa transcription : [http://research.cch.kcl.ac.uk/proust\\_prototype/about.html](http://research.cch.kcl.ac.uk/proust_prototype/about.html) et <http://goytisolo.unibe.ch/>



## 4. Conclusions et perspectives

Pour les trois années à venir (2014-2017), nos priorités seront donc de finir la description des fonds, la saisie des notices bibliographiques et l'édition de ces archives numériques sur un site performant accompagné d'un système permettant l'approche critique et génétique du manuscrit. Dans un premier temps, il est vrai que notre projet n'envisage pas d'études de lexicométrie ou de statistique textuelle, mais nous n'écartons pas le recours à des programmes comme la plateforme TXM<sup>45</sup> pour l'analyse de grands corpus de textes dans la mesure où nos outils et méthodes entrent en synergie avec les technologies de TXM.

Notre projet est à la croisée de différents domaines des SHS (littérature hispanique, critique génétique, théâtre, civilisation...), de l'archivistique, de l'informatique et du numérique, il est vrai qu'il évolue lentement car à chacune de ses étapes, la réflexion technique impose le temps de la discussion entre ses différents acteurs. Or c'est sans aucun doute l'un des principaux atouts du domaine nouveau qu'incarnent les *Humanités numériques*<sup>46</sup>: elles permettent d'interroger le monde dans lequel nous vivons, ses mutations et celles de la recherche à travers la réflexion et l'action collectives de personnes issues de domaines et métiers différents qui échangent des informations, des données et des recherches.

## Références

- Basso, E., C. Demasi, N. Dei Cas et F. Idmhand, (2012). Trayectoria de José Mora Guarnido. Espejo de un intelectual entre España y América (1923-1939). *Viajeros, diplomáticos y exiliados. Escritores hispanoamericanos en España* (1914-1939), C. De Mora et A. García Morales, Bruxelles, P.I.E. Peter Lang (2012:517-539).
- Barthes, R. (1993). *Le Bruissement de la langue*, Paris, Seuil, coll. Points Essais, 1993 (1993:63-69).
- Braillon-Chantraine, C. (2009). Fonds Carlos Denis Molina. *Catalogue des collections patrimoniales*, Université Lille 3 (2009:94-104).
- Burdick, A., J. Drucker, P. Lunenfeld, T. Presner, J. Schnapp (2012). *Digital Humanities*, Massachusetts, The MIT Press (2012:35).
- De Biasi, P-M (2003). *La génétique des textes*, Paris, Nathan PUF.

---

<sup>45</sup> Url : <http://textometrie.ens-lyon.fr/>

<sup>46</sup> « In contrast with most traditional forms of scholarship, Digital approaches are conspicuously collaborative and generative (...). This changes the culture of humanities work as well as the questions that can be asked of the materials and objects that comprise the humanistic corpus », in *Digital Humanities*, *op. cit.*, p.3.

- Douehi, M. (2008). *La grande conversion numérique*, Paris, Seuil (2008 :212).
- Idmhand, F. (2009). Le fonds José Mora Guarnido. *Catalogue des collections patrimoniales*, Université Lille 3 (2009:94-104).
- Idmhand, F. (2013). Échanges en exil, exils en échanges : la correspondance de José Mora Guarnido. *Migrations and exile: Charting New Literary and Artistic Territories*, A. Savin. Cambridge, Cambridge Scholars Publishing (2013 : 47-56).
- Vila Matas, E. (2010). *Dublinesca*, Barcelona, Seix Barral.

## Summary

Among the written heritage public or private institutions are likely to receive, the archive of contemporary writers are, by their original diversity and length, a challenge for those who are willing to receive, keep and preserve them and make them available. Regarding the Digital Humanities, they also consist in a major data bank which may be used and exchanged. The objective of the "Alcides Giraldi Fund" project is to take up the challenge and offer a complete "editorial method" for manuscripts.

This paper sets out to explain how, since 2003, we are engaged in the preservation of the "Alcides Giraldi." The fund, currently located in Lille, has many unpublished documents that will contribute to international research on the exilic movements and the circulation of ideas between Europe and the Americas during the major conflicts of the twentieth century.