



HAL
open science

CollEx-Persée Numérisation -Projet collaboratif Projet “ Archivos 3.0 ” Bilan scientifique et financier

Fatiha Idmhand

► **To cite this version:**

Fatiha Idmhand. CollEx-Persée Numérisation -Projet collaboratif Projet “ Archivos 3.0 ” Bilan scientifique et financier. [Rapport de recherche] CollEx Persée. 2021. halshs-03374612

HAL Id: halshs-03374612

<https://shs.hal.science/halshs-03374612v1>

Submitted on 12 Oct 2021

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.

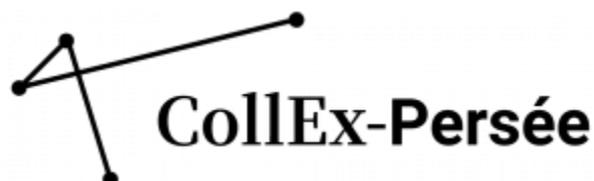
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

CollEx-Persée

Numérisation – Projet collaboratif

Projet « Archivos 3.0 »

Bilan scientifique et financier



Appel à projets septembre 2018 - Réalisation du projet 2019-2021

Rapport rédigé par Prof. Fatiha Idmhand (Université de Poitiers, ITEM UMR8132)

Contact : fatihaidmhand@yahoo.es / fatiha.idmhand@univ-poitiers.fr

Table des matières

1. Fiche d'identité du projet	3
2. Informations administratives	4
3. Bilan scientifique du projet	5
3.1. Les fonds d'archives du CRLA-Archivos	5
3.2. Un projet pour (re)créer des collections numériques et nourrir des études distantes sur les <i>médiateurs culturels</i>	5
3.2.1. Le traitement des fonds avant 2019-2021	6
3.2.2. Etat des lieux le 12-05-2021	7
3.2.3. La création et la gestion des données.....	8
3.3. Explorations scientifiques du projet.....	10
3.3.1. Qu'est-ce qu'un « médiateur culturel » ?	10
3.3.2. Quelle typologie pour les médiateurs culturels ?	11
3.3.3. Une « intelligence » des métadonnées ?	16
3.4. Conclusions : comment dégager des types ?	19
4. Calendrier du projet 2019-2021	20
5. Annexes.....	21
5.1.1. Plan de classement du Fonds Julio Cortázar	21
5.1.2. Plan de classement et inventaire du Fonds Fernando Aínsa (à la date du 12-05-2021).....	21
5.1.3. Plan de classement et inventaire du Fonds Jacques Gilard (à la date du 12-05-2021)	21

1. Fiche d'identité du projet

Thématique	Numérisation – Projet collaboratif
Titre du projet	Archivos 3.0
Établissement porteur du projet	Université de Poitiers – MSHS
Equipe de recherche concernée	CRLA-Archivos (Institut des Textes et Manuscrits Modernes, Poitiers, UMR 8132)
Corpus concerné	3 Fonds d'archives : <ul style="list-style-type: none"> - Fonds Julio Cortázar - Fonds Jacques Gilard - Fonds Fernando Aínsa
Indiquez si le corpus concerné est labellisé CollEx	Oui. Les corpus du CRLA-Archivos ont reçu la labellisation du CollEx en 2017
Actions à financer	x Numérisation X Production de métadonnées x Actions de valorisation, événements scientifiques x Développement technique d'un outil de production ou de diffusion existant
Durée prévisionnelle du projet	24 mois
Liste des partenaires	Institut des Textes et Manuscrits Modernes (UMR8132) Maison des Sciences de l'Homme et de la Société – Poitiers HumaNum Biblioteca Nacional del Uruguay (Montevideo, Uruguay) Universidad de la Plata (La Plata, Argentine) Casa de las Américas (La Havane, Cuba)
Résumé du projet (10-20 lignes max.)	Le projet Archivos 3.0 a trois objectifs principaux : - générer de nouvelles données numériques à partir des fonds d'archives qui ont été légués au CRLA-Archivos depuis 2009 et que la communauté scientifique attend impatiemment ; - organiser la reprise de données pour deux fonds d'archives qui ont été numérisés et mis en ligne avant 2005 et qui se trouvent actuellement édités dans des technologies obsolètes sans aucune interopérabilité ; - préparer ces corpus, essentiels pour la recherche scientifique dans le domaine latino-américain, au Web 3.0 dans le respect des principes FAIR et en vue du Web Sémantique. Le projet Archivos 3.0 s'inscrit résolument dans la perspective de l'Open Science.
Budget total du projet	131 600€
Financement demandé	62 500€

2. Informations administratives

Etablissement porteur du projet :

Personne ayant qualité pour engager l'établissement porteur du projet :

Nom	CHAUVEAU	Prénom	Frédéric
Fonction	Directeur MSHS – Gestion des crédits par la MSHS (USR du CNRS)		

Etablissement gestionnaire (si différent de l'établissement porteur du projet) :

Personne ayant qualité pour engager l'établissement porteur du projet :

Nom	CHAUVEAU	Prénom	Frédéric
Fonction	Directeur MSHS – Gestion des crédits par la MSHS (USR du CNRS)		

Etablissement qui conserve le corpus traité (si différent de l'établissement porteur du projet) :

Personne ayant qualité pour engager l'établissement :

Nom	CRLA Archivos	Prénom	
Fonction	Université de Poitiers CNRS (UMR 8132)		

Autre partenaire(s) du projet :

Personne ayant qualité pour engager chaque partenaire du projet :

Nom	DIORIO	Prénom	Paolo
Fonction	Directeur de l'Institut des Textes et Manuscrits Modernes (UMR8132)		

Personne responsable du projet (interlocuteur privilégié de Collex-Persée tout au long de la procédure d'appel à projets, il relaiera les informations auprès des différents partenaires)

Nom : IDMHAND, Fatiha

Fonction : Professeur

Adresse : Université de Poitiers

Courrier électronique : fatihaidmhand@yahoo.es ou fatiha.idmhand@univ-poitiers.fr

3. Bilan scientifique du projet

3.1. Les fonds d'archives du CRLA-Archivos

Les archives de l'équipe de l'Institut des Textes et Manuscrits Modernes (UMR8132) « CRLA-Archivos » constituent l'un des plus importants fonds d'archives et manuscrits d'écrivains latino-américains de France. Il a été constitué à partir de legs successifs qui ont été réalisés, depuis 1984, par les auteurs et intellectuels eux-mêmes ou par leurs ayants-droits.

Située à l'Université de Poitiers, le « Centre de Recherches Latino-Américaines-Archivos » est une équipe de l'Institut des Textes et Manuscrits Modernes –ITEM (UMR8132). Elle est née en 1995, avec la fusion du Centre de Recherches Latino-Américaines de l'Université de Poitiers (1964-1995) et du GDR 1190 « Archives de la Littérature Latino-Américaine et des Caraïbes du XXe siècle » (CNRS-Université de Paris X-Nanterre). Ce GDR avait porté le projet de création de la collection « Archivos » en 1984 avec le soutien de l'UNESCO (<http://crla-archivos.labo.univ-poitiers.fr/la-collection-archivos/>). Dépendant de l'Institut des Textes et Manuscrits Modernes (UMR8132), l'équipe du « CRLA-Archivos » est localisée à Poitiers, ses bureaux et fonds d'archives sont hébergés par la Maison des Sciences de l'Homme et de la Société de l'Université de Poitiers.

La MSHS de Poitiers est une Unité de Services et de Recherche du CNRS (USR 3565) qui accompagne les laboratoires de recherches en leur offrant des services (locaux, accompagnements administratifs, etc.) et en mettant à leur disposition des équipements informatiques spécifiques et notamment une plateforme technologique (Numérisation, Géomatique, Analyse du comportement Humain en temps réel [EEG], oculométrie). C'est au sein de cette plateforme que la numérisation des archives du « CRLA-Archivos » est réalisée.

3.2. Un projet pour (re)créer des collections numériques et nourrir des études distantes sur les *médiateurs culturels*

Le projet « Archivos 3.0 » présenté lors de l'appel à projet 2018 s'était donné deux objectifs principaux ; le premier visait à organiser la reprise de données de fonds d'archives qui avaient déjà été numérisés par le passé mais dont les données n'étaient plus ou difficilement accessibles comme le fonds Julio Cortázar. Le but était ainsi de permettre la recherche scientifique sur ces fonds depuis n'importe quel pays du monde. Le second objectif visait à accélérer le travail entrepris sur deux nouveaux fonds d'archives : les fonds Jacques Gilard et Fernando Aínsa. Il est apparu que ces fonds pouvaient, par ailleurs, contribuer à développer les études *distantes*¹ sur la circulation des idées et des œuvres entre l'Europe et l'Amérique latine durant les principaux conflits du XX^e siècle ; « Archivos 3.0 » s'est donc donné pour objectif de construire et de livrer à la communauté internationale un corpus de données et de métadonnées sur les médiateurs culturels et de tester des méthodes informatiques permettant d'explorer ces données.

Les corpus choisis pour la proposition étaient les archives de Julio Cortázar, de Jacques Gilard et de Fernando Aínsa, trois corpus sélectionnés parce qu'ils apportent des éclairages importants sur la circulation de la littérature latino-américaine entre le continent américain et l'Europe, sur sa réception en France et sur le rôle de certains chercheurs français en tant que *médiateurs* des transferts culturels :

¹ Le concept de « *distant reading* » a été imaginé par Franco Moretti (2003) dans son article « Conjectures on world literature (*Distant reading*, 2013). Au cours des dernières années, le « *distant reading* » (expression pour laquelle un équivalent consensuel ne semble pas encore avoir été trouvé en français) s'est imposé comme l'une des méthodologies les plus prometteuses pour les études littéraires assistées par ordinateur. Voir Galleron, Idmhand, Meynard in « Que mille lectures s'épanouissent » in Digital Humanities Quarterly (DHQ) <http://www.digitalhumanities.org/dhq/preview/index.html>

- **Julio Florencio Cortázar Descotte (1914-1984)** est né le 26 août 1914 à Ixelles (Belgique), il est mort le 12 février 1984 à Paris. Argentin, il a grandi à Buenos Aires avant de partir s'installer à Paris en 1951 où il a résidé jusqu'à son décès en 1984. Auteur de romans et de nouvelles en langue espagnole, il a été lauréat de très nombreux prix : *Marelle* (1963) est son roman le plus commenté tandis que *Le livre de Manuel* (1973) a reçu le prix Médicis du roman étranger en 1974. Écrivain engagé, il s'est impliqué dans la défense des droits de l'homme et s'est publiquement opposé au gouvernement de Juan Domingo Perón.
- **Jacques Gilard (1943-2008)** a été Maître de conférences à l'Université de Toulouse Le Mirail et Directeur de l'Institut pluridisciplinaire pour les études sur l'Amérique latine à Toulouse (IPEALT). Eminent spécialiste des études caribéennes en France, il a consacré sa carrière à la littérature latino-américaine, notamment à l'œuvre du colombien Gabriel García Márquez. Directeur des revues « Caravelle » et « L'Ordinaire Latino-américain », il a personnellement connu Alejo Carpentier et José Lezama Lima. Ses archives comportent une collection et une documentation étonnantes issues d'archives de presse et de bibliothèques publiques, familiales et privées de Barranquilla.
- **Fernando Aínsa (1937-2019)** était un écrivain uruguayen d'origine espagnole. Il a travaillé à l'Unesco (Paris) de 1974 à 1999 où il a été directeur littéraire des Éditions UNESCO. Critique, spécialiste de la littérature et de la pensée hispano-américaine, il a développé des travaux internationalement reconnus sur les concepts d'identité et d'utopie. Il a également été auteur de romans, de recueils de nouvelles et de poèmes. Ses archives illustrent de quelles façons son poste à la direction des Éditions UNESCO à Paris lui a permis de se situer au centre des réseaux d'échanges intellectuels et académiques entre les Amériques et l'Europe.

Ces trois corpus permettent d'aborder le champ latino-américain depuis la Caraïbe jusqu'au Río de la Plata, alors que leur intérêt scientifique est avéré depuis de nombreuses années, l'impact consécutif à la mise à disposition numérique de ces archives nous est apparu prometteur.

3.2.1. Le traitement des fonds avant 2019-2021

Avant 2018, seul le fonds Julio Cortázar avait été inventorié, organisé et numérisé. Les archives virtuelles étaient consultables en ligne sur un site web (<http://www.mshs.univ-poitiers.fr/crla/contenidos/Cortazar/>) sur lequel avaient été publiées les images numériques des documents sous forme de miniatures dont la qualité avait été dégradée pour que la visionneuse permette l'affichage sur le web. Ces images étaient accompagnées de métadonnées. En raison de l'obsolescence de la plupart des technologies utilisées pour construire le site web et de l'absence de maintien technique, l'accès à l'information était compromis. « Archivos 3.0 » s'est proposé de reprendre ces données (accessibles malgré tout dans une base de données Php-MySQL) et de les réinsérer dans un nouveau cycle numérique.

Dans le cas des archives de Jacques Gilard et de Fernando Aínsa : tout était à faire. Fernando Aínsa avait légué ses archives au CRLA-Archivos peu avant l'année 2016 mais le travail de classement, d'inventaire et de numérisation n'a pu commencer qu'en 2018 grâce au soutien de l'Institut des Textes et Manuscrits Modernes à travers le projet « Labex TransferS ». La correspondance de cet auteur a été traitée en premier lieu : 1794 lettres purent être décrites, numérisées et mises en ligne sur : <http://ainsa.nakalona.fr/>.

Enfin, dans le cas du fonds Jacques Gilard, quelques explorations ont été entreprises en 2017, mais de façon très partielles puisque le fonds n'est pas classé et est encore dispersé dans plusieurs cartons. Le pré triage réalisé avait estimé le corpus à au moins une quinzaine de milliers de documents.

C'est donc une révision totale du processus de traitement des fonds d'archives qui a été mis en place grâce au projet « Archivos 3.0 », un processus tient compte du cycle de vie des données et qui s'appuie sur les outils de l'infrastructure numérique Huma-Num pour traiter, stocker et diffuser les données produites.

3.2.2. Etat des lieux le 12-05-2021

a) Le fonds Julio Cortázar

Le projet « Archivos 3.0 » s'est appuyé sur les travaux qui avaient été réalisés par l'équipe « Archivos » avant 2016 afin de préserver le référencement et la citation de ce fonds: le plan de classement par « Carpeta » n'a pas été modifié et les regroupements d'origine conservés. Le travail a consisté à réviser et enrichir l'existant selon une approche systématique :

- tous les documents ont été contrôlés et inventoriés un à un de même que toutes les images numériques : les numérisations manquantes ont été réalisées et les images de mauvaise qualité reprises de façon à remplacer l'existant le cas échéant ;
- les métadonnées publiées sur le site <http://www.mshs.univ-poitiers.fr/crla/contenidos/Cortazar/> ont été extraites depuis la base de données Php-MySQL et réorganisées (à l'aide de l'outil OpenRefine). Un nouveau fichier de données a été produit (tableur) et les informations révisées et actualisées. Cette étape d'enrichissement des métadonnées avait pour but d'assurer la présence, pour chaque document, d'un minimum d'informations et de les normaliser selon les recommandations des standards internationaux, comme le Dublin Core. Une comparaison entre les informations exposées sur le site web d'origine et celles qui peuvent être consultées sur l'actuel site web (<https://cortazar.nakalona.fr>) témoigne de l'important travail qui a été réalisé ;
- enfin, la collection a été enrichie de nouveaux documents issus d'un legs réalisé durant l'année 2018 « Donación Yurkievich ». Ces documents ont été numérisés et intégrés aux archives numériques de façon à compléter la collection accessible en ligne.

- **A cette date, le traitement de ce fonds est terminé : 2403 documents ont été repris et 18736 fichiers images ont été produits (62,5 Go en Jpeg).**

b) Le fonds Fernando Aínsa

Arrivé à Poitiers un peu avant l'année 2016, un premier travail de classement de ce fonds a pu être réalisé durant le printemps 2017 grâce au soutien du « Labex TransferS » (<http://www.transfers.ens.fr/>). Cet examen préliminaire a permis de repérer et d'estimer à près de 4000 lettres le corpus des correspondances.

L'objectif du projet Archivos 3.0 a été de poursuivre la constitution de ce nouveau corpus en décrivant, à la pièce, chacun des documents de ce fonds d'archives et prioritairement, les correspondances. Ce fonds est encore inconnu de la communauté scientifique, mais la publication de ces données constituera une importante innovation. L'actuel site web permet de consulter une partie des données décrivant les documents (<https://ainsa.nakalona.fr>).

- **A la date de rédaction de ce rapport, le classement de ce fonds n'est pas terminé. A ce stade, seule la correspondance a été traitée et le corpus épistolaire représente pour l'instant 4113 documents décrits, 6225 images (2,61 Go en Jpeg).**
- Une thèse doctorale est actuellement en cours d'élaboration sur ces archives², elle s'appuie sur les données issues de ce fonds d'archives et vise une étude comparative sur le profil de médiateurs culturels de deux intellectuels qui ont travaillé à l'Unesco : Roger Caillois et Fernando Aínsa.

² Thèse actuellement réalisée par Yadira Carranza Marino (2019-2022). Le titre est : « Fernando Aínsa et Roger Caillois: Médiateurs des transferts culturels entre l'Amérique latine et l'Europe »

c) Le fonds Jacques Gilard

Le fonds Jacques Gilard est celui sur lequel l'équipe a le moins avancé pour deux raisons. La première concerne sa situation physique : les documents sont encore très désorganisés et dispersés dans une quarantaine de cartons. Plusieurs groupes d'étudiants, à l'occasion de courts stages de 6 semaines³, nous ont aidés à pré-classer les documents par type (correspondances, presse, manuscrits, etc.) mais cette tâche, longue et difficile à mener, a été impactée par la crise Covid en 2020.

- **A la date de rédaction de ce rapport, le classement du fonds n'est pas terminé. Ce corpus, estimé à au moins vingt mille documents, ne pourra pas être traité dans son intégralité, mais l'objectif est de proposer quelques dossiers thématiques et quelques correspondances. Seuls 859 documents ont été décrits à ce stade et 4936 images numérisées (13,3 Go en Jpeg) et 59 fichiers textes extraits de disquettes. Ils concernent le dossier thématique « Barranquilla »**
- Dans le cadre d'un Master 2, ce dossier thématique a pu être constitué et l'étude réalisée a confirmé l'intérêt du dossier. Celui-ci porte sur la naissance du groupe de « Barranquilla », sur le rôle de Gabriel García Márquez au sein de celui-ci et notamment l'impact du groupe dans l'émergence de ses idées novatrices.

3.2.3. La création et la gestion des données

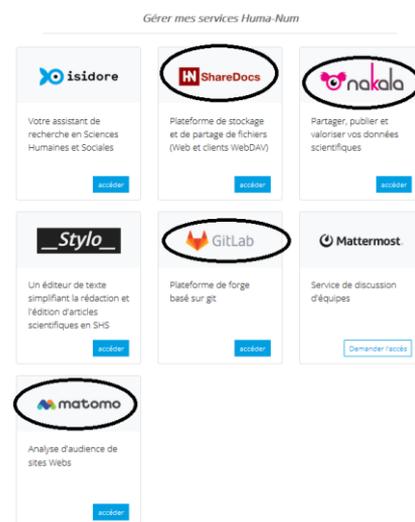
Comme indiqué précédemment, l'équipe du projet utilise les services que l'infrastructure Huma-Num met à disposition des équipes scientifiques pour gérer et stocker les données constituées⁴. Les services suivants sont tout particulièrement utilisés : Sharedocs pour le travail quotidien sur les données, Nakala pour stocker les données en vue de leur archivage à long terme, Gitlab pour déposer les codes informatiques des outils développés et Matomo pour suivre l'audience des sites webs développés.

Les équipements fournis par la MSHS sont utilisés pour « capter » les données et notamment pour numériser les archives.

C'est surtout pour accompagner cette étape du projet qu'un ingénieur-archiviste a été recruté par « Archivos 3.0 » :

a) Tri et classement des archives

Cette tâche est une étape longue et chronophage du travail sur les archives et manuscrits. Lorsque les fonds n'ont jamais été triés, elle est éminemment coûteuse en temps mais indispensable. Dans le cas du projet « Archivos 3.0 », le tri peut être réalisé dans les locaux de l'équipe « CRLA-Archivos » situés au premier étage de la MSHS. Durant la pandémie Covid de 2020, les restrictions d'accès ont parfois compromis cette tâche qui ne peut être réalisée qu'en présentiel. Les explorations initiées en vue du classement du fonds Jacques Gilard ont été arrêtées net alors qu'une étudiante de Master 2 était engagée dans une recherche sur ce fonds d'archives et lorsque l'accès aux locaux a de nouveau été possible, nous avons décidé d'intensifier le travail de description du Fonds Fernando Aínsa. Globalement, lorsque les fonds ne sont pas triés, le



³ S'agissant de stages de moins de huit semaines, ils ont donné lieu, le plus souvent, à des gratifications sous forme de bons d'achats.

⁴ <https://www.huma-num.fr/les-services-par-etapes/>

classement est organisé en deux étapes : un pré triage permet de comprendre l'organisation du fonds et de construire un projet de plan de classement. La description de tous les documents à la pièce permet ensuite de recenser toutes les pièces de façon exhaustive, le plan est alors confirmé ou modifié, et publié accompagné de l'inventaire détaillé du fonds comme l'illustrent les trois annexes jointes à ce rapport.

Concernant les fonds déjà triés et classés, comme le fonds Julio Cortazar, cette étape consiste en une révision complète des documents à des fins de contrôle (repérage d'éventuels oublis), de correction ou d'actualisation des informations.

b) Numérisation

La numérisation est réalisée à la MSHS à l'aide de scanners de différentes tailles. Quelques métadonnées sont pré-saisies de façon à gagner du temps lors du traitement des images. Les images collectées sont immédiatement contrôlées et renommées de façon à correspondre aux métadonnées descriptives. Elles restent stockées sur les serveurs de la MSHS et, pour faciliter le travail en équipe, elles sont également stockées sur le « cloud » de l'équipe grâce à Sharedocs. Des copies des images sont également réalisées sur différents disques durs externes.

c) Saisie des métadonnées

La description de chaque document tente d'être aussi exhaustive que possible, elle est réalisée parallèlement à la numérisation. Cette troisième tâche, manuelle et chronophage, peut se révéler assez longue car, outre la description physique du document, nous apportons une série d'informations contextuelles et un résumé qui permettent à l'internaute de prendre connaissance du contenu intellectuel. La recherche d'information peut se révéler assez longue lorsqu'il s'agit de retrouver des informations sur l'auteur d'une lettre ou sur la provenance d'un extrait de coupure de presse amputé du nom du journal et/ou de la date.

La saisie des métadonnées repose sur un modèle qui a été élaboré en amont, avec le chercheur⁵ : cette tâche est essentielle car elle facilite le travail scientifique. Pour la réaliser, aucun outil de gestion spécifique n'a été développé à cette fin, le tableur est l'instrument privilégié par le projet, il est facile à utiliser et peut être aisément adapté lors d'explorations informatiques.

d) Exposition et stockage des données

Ces deux tâches représentent deux étapes qui sont partiellement finalisées à ce stade. Dans la mesure où le projet « Archivos 3.0 (2) » se poursuit, les deux tâches seront finalisées en 2022. Toutefois, les principes retenus seront les suivants :

- Le stockage des données à des fins d'archivage à long terme sera réalisé grâce à l'outil « Nakala » d'Huma-Num (<https://nakala.fr/>). Il permet de préserver le travail réalisé et de rendre les données FAIR. L'infrastructure Huma-Num n'a livré la version finale de cet outil qu'en décembre 2020. Parallèlement, le consortium CAHIER a développé une application de dépôt en lot qui sera disponible à la fin du printemps 2021. L'application *mynakala*⁶ sera utilisée pour rendre les données du projet CollEx « Archivos 3.0 » FAIR.
- Pour exposer les premières données collectées, l'outil « Nakalona » de Huma-Num avait été utilisé pour concevoir deux bibliothèques virtuelles (<https://cortazar.nakalona.fr/> et <https://ainsa.nakalona.fr/>) mais Huma-Num va cesser de proposer Nakalona et a élaboré un nouvel outil de visualisation qui s'appuie sur « Nakala ». Grâce à « Nakala Press », qui sera présenté fin mai 2021, il sera possible d'exposer et d'éditorialiser les données stockées sur Nakala. Cet outil n'est pas encore disponible à l'heure où ce rapport est rédigé. La stratégie

⁵ Dans le cas présent, le modèle a été développé en 2014-2017, dans le cadre du projet ANR-CHispa (ANR-13-JSH3-0006). Le rapport de fin de projet est accessible sur HAL : <https://halshs.archives-ouvertes.fr/halshs-01899897> (il a été publié en quatre parties). Voir également le blog : <https://chispa.hypotheses.org/>

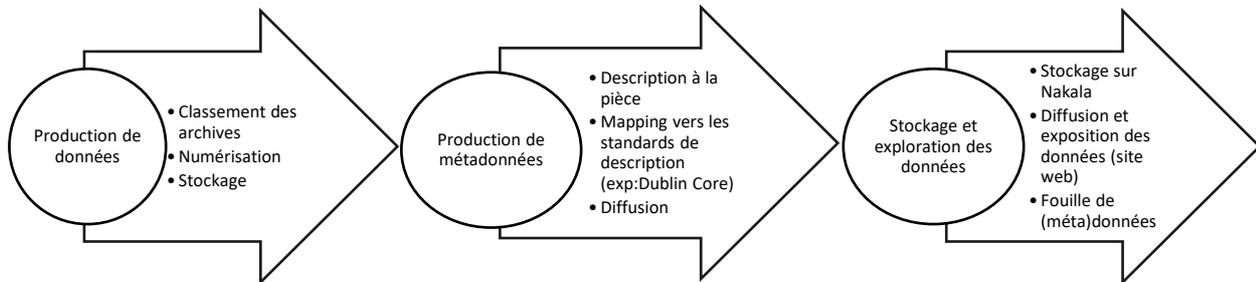
⁶ <http://mynakala.huma-num.fr/>

envisagée par le projet consistera à utiliser « Nakala press » et/ou un autre service d'exposition de données comme celui qui est proposé à l'Université de Poitiers par les collègues de l'IST à l'aide du CMS Omeka : le service BivUp⁷.

Ces solutions seront mises en œuvre pour la fin du projet « Archivos 3.0 (2) » en 2022.

e) Méthode de travail

La méthode de travail de l'équipe peut être résumée en trois volets comme illustré ci-dessous :



3.3. Explorations scientifiques du projet

3.3.1. Qu'est-ce qu'un « médiateur culturel » ?

L'un des objectifs de la dématérialisation des archives de l'équipe « CRLA-Archivos » est de contribuer à construire un nouveau corpus numérique (données et métadonnées) permettant d'étudier les « transferts culturels ». C'est à Michel Espagne que nous devons la création, dans les années 1980⁸, de la notion de « transferts culturels », le concept avait pour but de poser un nouveau cadre méthodologique pour les projets de recherche « trans-nationaux ».

La notion de « transferts » vise l'étude des processus de sélection et de « resémantisation » qui accompagnent la transmission des objets culturels et des idées entre les aires culturelles. La notion de « transfert culturel » met ainsi en perspective la circulation des objets d'une part et les modalités de leurs importations dans les nouveaux champs culturels de l'autre. Ces importations sont examinées à l'aune des réinterprétations auxquelles elles donnent lieu dans le nouveau contexte, de leurs redéfinitions, et des processus qui accompagnent leurs re-significations.

Ainsi que l'a spécifié Michel Espagne, les « transferts culturels » sont facilités par l'action concrète de celles et ceux qui agissent comme des passeurs, comme les « vecteurs » de ces transferts⁹ : il s'agit des intellectuels, traducteurs, exilés, expatriés ou universitaires qui facilitent les transferts ; Michel Espagne les a appelés les « médiateurs des transferts culturels ». Cependant, il n'existe à ce jour aucune typologie de leurs actions, aucune catégorie permettant de décrire et de classer les médiateurs en fonction de ces mêmes actions. Ces dernières années, la création d'une telle typologie nous a semblée pertinente pour l'étude de ces médiateurs : un tel outil permettrait de mieux comprendre leur rôle dans la circulation des idées voire, de les identifier et même de les classer. L'existence d'une typologie permettrait en outre, de sortir d'une certaine « personnalisation » de la circulation des idées ; en effet, la critique tend parfois à réduire l'introduction des idées ou la genèse de mouvements de pensées au nom d'un(e) seul(e) « chef(fe) de file » ou à une poignée de personnalités. Une autre approche de la circulation des idées serait donc possible si nous nous

⁷ <http://bivup.univ-poitiers.fr/>

⁸ Michel Espagne, « La notion de transfert culturel », Revue Sciences/Lettres [En ligne], 1 | 2013, mis en ligne le 01 mai 2012. URL : <http://rsl.revues.org/219> ; DOI : 10.4000/rsl.219

⁹ Michel Espagne, « Transferts culturels - Pays germaniques », in <http://www.umar8547.ens.fr/spip.php?rubrique7>

intéressions aux réseaux des figures de proue, aux personnes qui les ont entourées et qui ont accompagné la genèse de leurs projets et leurs diffusions.

En 2014, nous avons proposé d'introduire une première catégorie dans le classement de ces « médiateurs », il s'agit de la notion d'intellectuels « satellites »¹⁰. Ce concept désigne un groupe spécifique de *médiateurs* : celles et ceux qui appartiennent au réseau d'une « star » et qui se situent à moins d'un degré de distance de celles-ci si l'on se fonde sur la théorie des réseaux. Il s'agit donc du tout premier cercle de proches, des membres de la famille ou des amis qui, depuis l'ombre le plus souvent, contribuent à la construction de la renommée internationale des stars en accompagnant la genèse de leurs œuvres ou leur circulation à travers les aires culturelles. La question est donc de savoir s'il est possible de lister les actions qu'ils mettent en place à cette fin et s'ils existent des similitudes entre les médiateurs. Pour répondre à cette question, leurs archives apparaissent comme un réservoir d'informations privilégié et à cette fin, les métadonnées qui les décrivent se présentent comme des sources à exploiter.

3.3.2. Quelle typologie pour les médiateurs culturels ?

Pour construire notre typologie, nous nous sommes proposés d'utiliser les métadonnées des fonds d'archives comme moyens d'identification des propriétés et traits communs. La liste et le classement de ces propriétés et traits permettront, dans un second temps, de classer les médiateurs des transferts culturels en fonction de leurs rôles et actions. L'expérience menée par le projet « Archivos 3.0 » (2018-2020) a consisté à imaginer des méthodes de fouille de métadonnées.

a) Le jeu de métadonnées

Pour décrire nos fonds d'archives d'auteurs et intellectuels, nous utilisons un jeu de métadonnées aussi exhaustif que possible¹¹ : le but recherché est la construction de la description la plus complète du document afin que l'internaute qui consulte les métadonnées en ligne, puisse prendre connaissance du contenu de la ressource, en comprendre le fond, la forme et les enjeux. Notre jeu de métadonnées comporte une série de 46 étiquettes, elles sont adaptées en fonction des types de documents décrits (presse, œuvre de l'esprit, correspondance, etc.) et des besoins de compréhension du document, des informations disponibles sur sa genèse ou sur ses liens avec d'autres ressources du fonds d'archives décrit ou d'ailleurs (Sujet, Description, Notes, Résumé, État génétique, Relations génétiques, Autres ressources en relations). Le jeu de métadonnées comporte les informations suivantes :

Cote, Type, Date, Titre, Sous-titre, Auteur, Destinataire (lorsqu'il s'agit de correspondances), Sujet, Description, Notes, Résumé, Responsable scientifique, Éditeur scientifique, Responsable du fonds, Éditeur, Localisation, Droits, Licence, Ayants droit, Format, Langue, État génétique, Relations génétiques, Autres ressources en relations, Nature du document, Support, État général, Publication, Représentation, Contexte géographique, Lieu d'expédition (lorsqu'il s'agit de correspondances), Type de publication, Titre de la publication, Lieu de publication, Numéro de la publication (lorsqu'il s'agit de revues), Périodicité (lorsqu'il s'agit de revues), Directeur de la publication (lorsqu'il s'agit de revues), Auteur de l'analyse, Date de l'analyse, Auteur de la notice, Date de création de la notice, Auteur de la révision, Date de la révision de la notice, Auteur transcription, Donnée publique/Donnée privée.

¹⁰ <http://revistaseug.ugr.es/index.php/letral/article/view/6665>

¹¹ Voir supra note n°5 : le modèle a été développé en 2014-2017, dans le cadre du projet ANR-CHispa (ANR-13-JSH3-0006).

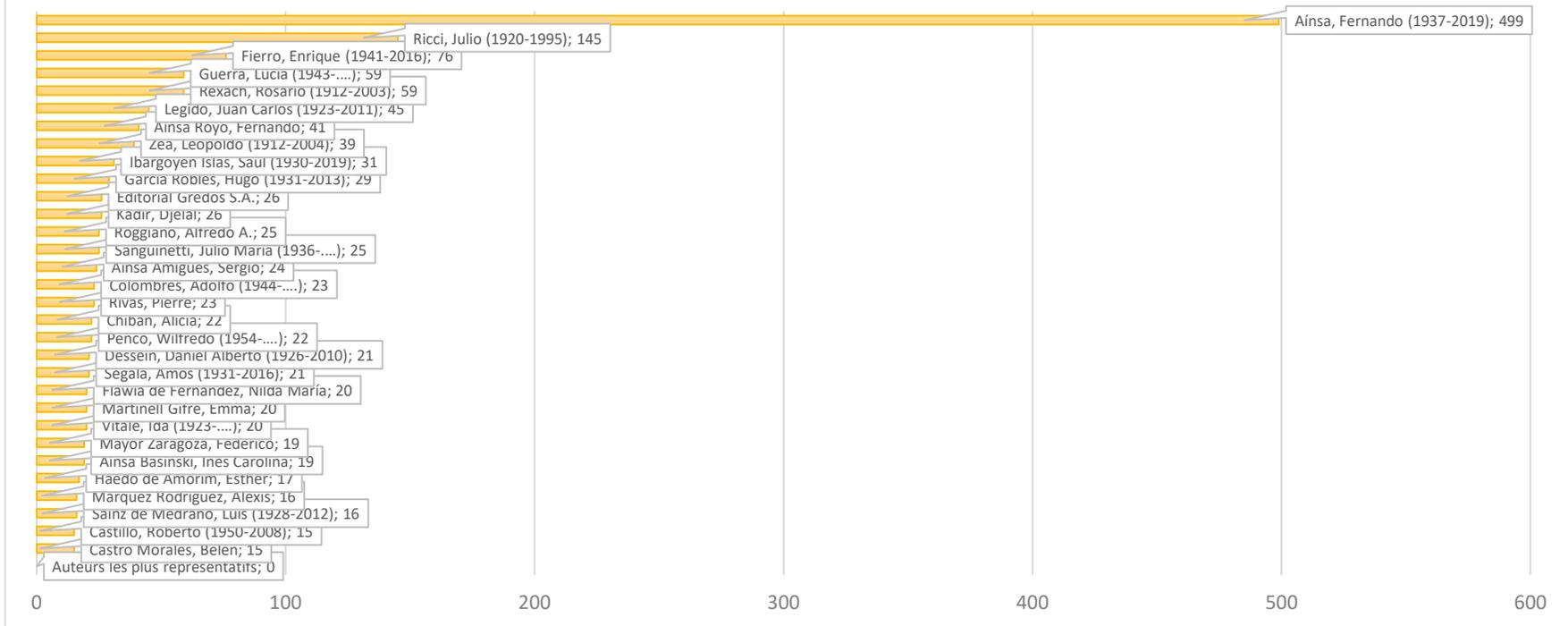
Lors de la mise en ligne des documents numérisés (ou lors du dépôt des données à des fins de stockage ou de diffusion sur le web), les métadonnées du projet sont ensuite associées via un processus de *mapping* aux étiquettes attendues par l'outil d'exposition, de stockage ou d'édition (Dublin Core ou XML-TEI ou XML-EAD, etc.).

Dans le cas du Dublin Core, ou, plus généralement, du dcterms. Cette opération s'effectue sans perte d'information puisque nous tirons le meilleur profit des champs répétables du Dublin Core en réalisant les *mappings* suivants :

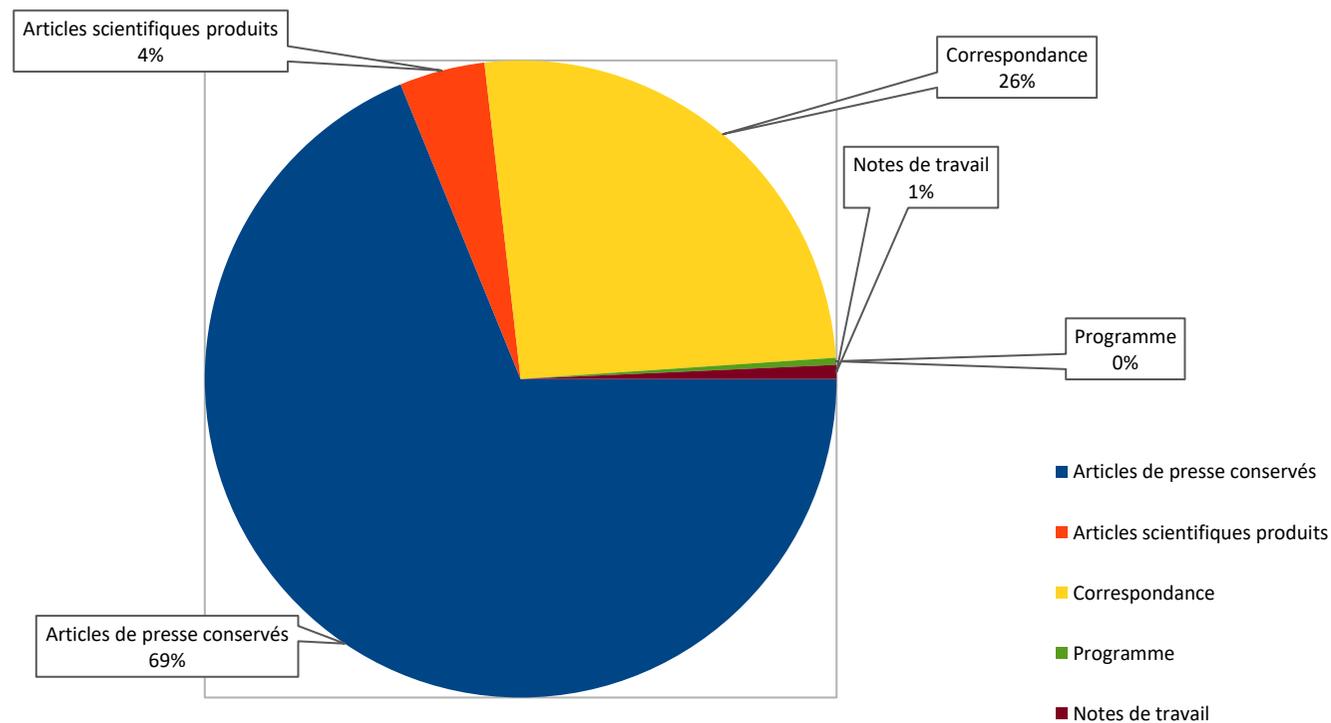
Métadonnées préparées	Métadonnées du Dublin Core (dcterms)
Titre	dc-rdf:title
Sous-titre	
Sujet	dc-rdf:subject
Créateur	dc-rdf:creator
Source	dc-rdf:source
Support	dc-rdf:medium
Éditeur	dc-rdf:publisher
Contributeur	dc-rdf:contributor
Langue	dc-rdf:language
Format	dc-rdf:format
Nature du document	
État général	
Identifiant	dc-rdf:identifier
Résumé	dc-rdf:abstract
Date de création	dc-rdf:created
Date de modification	dc-rdf:modified
Date de disponibilité	dc-rdf:available
Couverture spatiale	dc-rdf:spatial
Contexte géographique	
Lieu d'expédition	
Lieu de publication	
Droits	dc-rdf:rights
Licence CC	
Ayants droit	dc-rdf:rightsHolder
Référence bibliographique	dc-rdf:bibliographicCitation
Description	dc-rdf:description
État génétique	
Notes	
Type de publication	
Représentation	
Périodicité	
Numéro de la publication	
Directeur de la publication	
Relation	
Autres ressources en relations	
Est une version de	dc-rdf:isVersionOf

A partir de ces riches métadonnées, différentes visualisations peuvent être réalisées. Les images suivantes permettent de prendre connaissance des « profils » des fonds d'archives et de leurs auteurs. Les vues créées concernent 859 documents actuellement décrits du fonds Jacques Gilard, 4113 correspondance du fonds Fernando Aínsa et la totalité des 2403 documents qui composent le fonds Julio Cortázar :

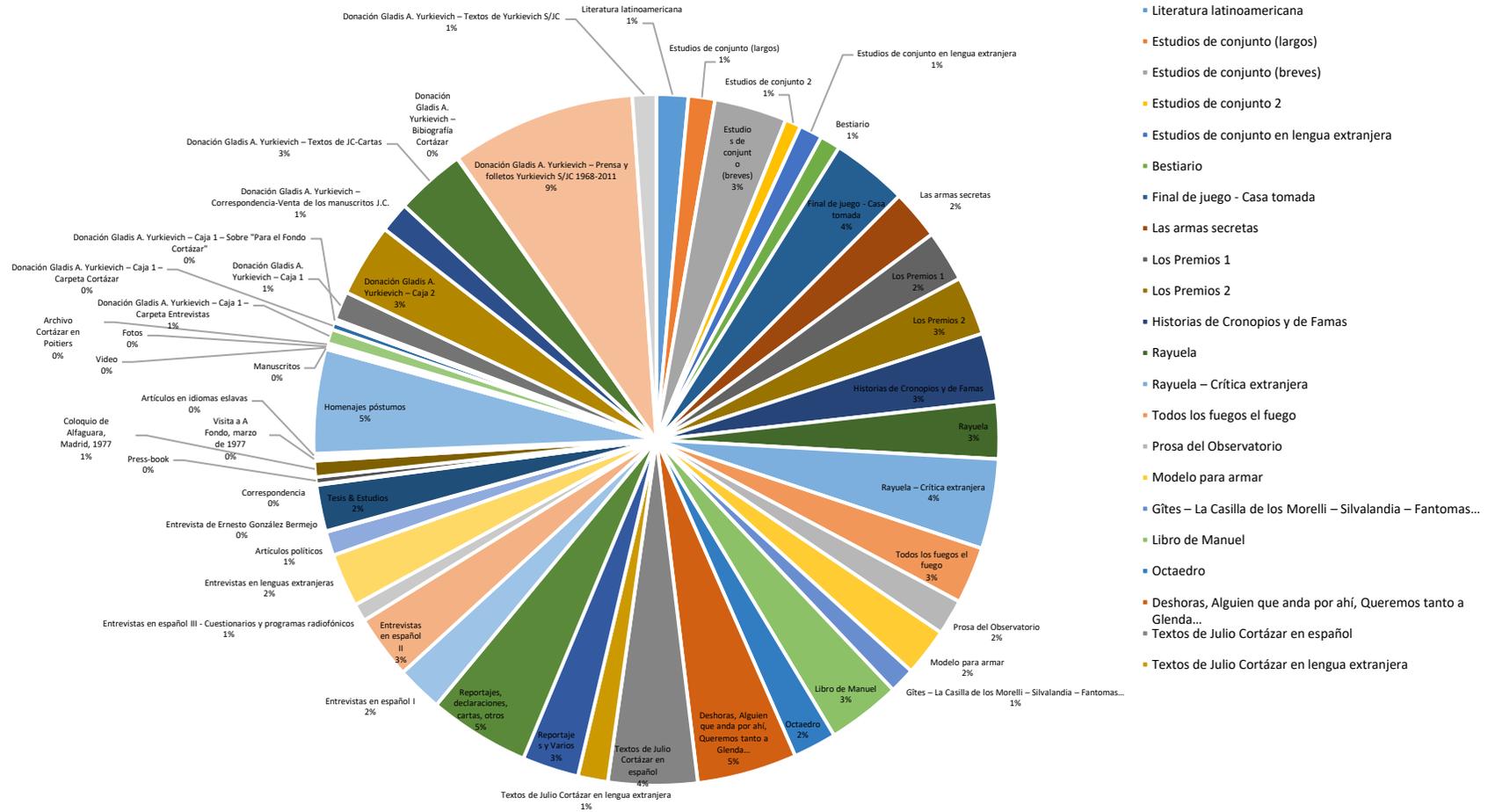
Correspondants de Fernando Aínsa (classés par nombre de lettres)



Types de documents indexés dans le fonds Jacques Gilard (à la date du 12-05-2021)

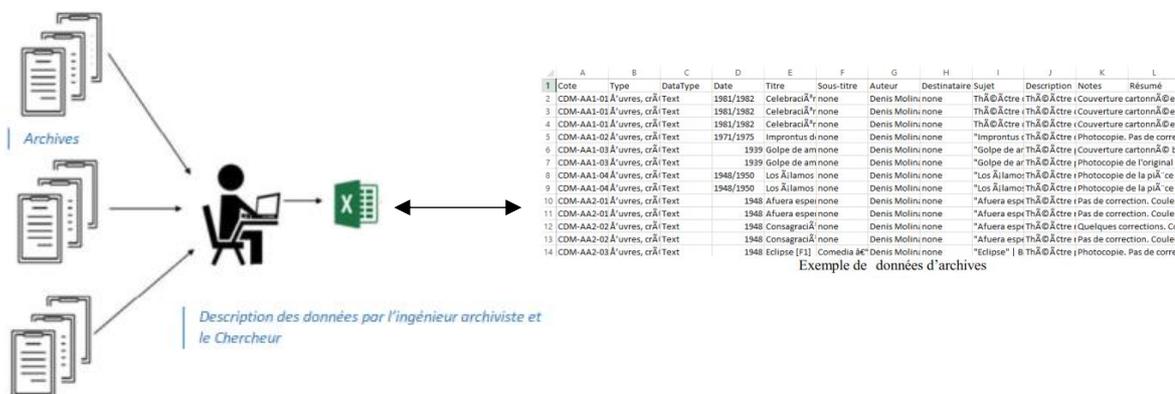


Types of documents indexés fonds Julio Cortázar



b) Comment identifier les « traits communs » ?

L'hypothèse testée a consisté à utiliser les métadonnées des documents des fonds d'archives, pour tenter d'identifier des traits communs et esquisser les profils des médiateurs. Ainsi par exemple, nous avons constaté que tous les fonds d'archives dits « d'intellectuels » du XX^e siècle comportaient des correspondances et de la presse (soit parce que les auteurs ont écrit dans la presse, soit parce qu'ils ont été interviewés, soit parce qu'ils ont conservé des articles sur des sujets qui les intéressent) mais que seuls les fonds de ceux qui ont eu une activité de création (comme Julio Cortázar et Fernando Aínsa) comportaient des manuscrits de projets inventifs. Toutefois, on peut trouver, comme dans le fonds Jacques Gilard, des manuscrits de projets inventifs qui concernent d'autres auteurs. Une fouille des métadonnées fondée sur toutes ou parties de nos étiquettes permettrait-elle de dessiner le portrait-robot de nos médiateurs ? C'est l'hypothèque que nous avons étudiée en abordant les métadonnées saisies sous une même étiquette comme autant de propriétés susceptibles de faciliter le repérage de traits communs ou de réaliser des comparaisons massives.



A partir des tableurs constitués par l'ingénieur-archiviste, des expériences de regroupements (*clustering*) ont été mises en place. Trois points de vues ont été explorés : la chronologie, la langue et les thématiques (mots clés) à l'aide des étiquettes suivantes : « date », « subject », « language » et, dans le cas des correspondances, « destinataire » et « creator ». A partir de celles-ci, des méthodes d'analyses informatiques ont été expérimentées grâce à une association avec le Laboratoire LIAS (Laboratoire d'Informatique et d'Automatique pour les Systèmes¹²) et au recrutement de deux stagiaires en Master 2 d'informatique en 2019 et 2020.

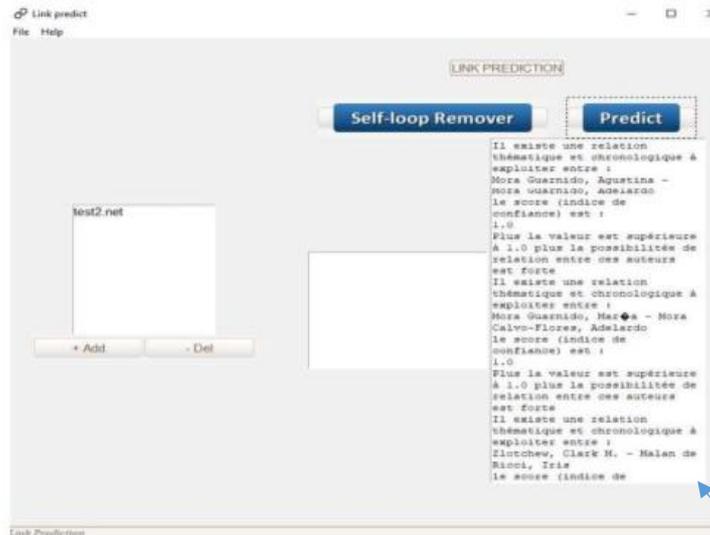
3.3.3. Une « intelligence » des métadonnées ?

a) Le développement de l'outil « Link&Pred »

Le recrutement d'un premier stagiaire en 2019 a permis de réfléchir à l'application de différentes méthodes d'extraction de connaissances (ECD) à partir des métadonnées. Un état des lieux des algorithmes d'apprentissages supervisés (Machine Learning) a permis de sélectionner ceux qui seraient adaptés aux caractéristiques de nos bases de données qui décrivent des corpus d'auteurs. L'idée retenue a consisté à tester les méthodes de prédictions de liens sur la base d'attributs simples,

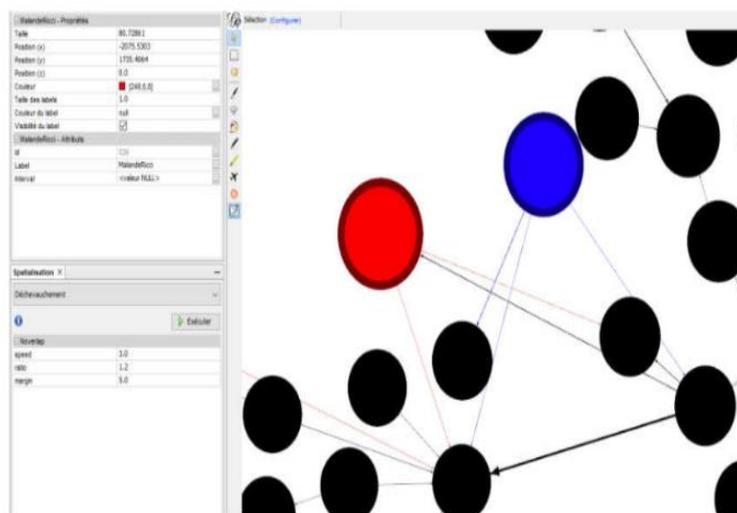
¹² <https://www.lias-lab.fr/>

pour cela, un prototype d'outil a été conçu, « Link&Pred » (voir image ci-dessous)¹³, pour accompagner la fouille de métadonnées soutenue par les méthodes actuelles de l'apprentissage automatique.



Résultats de l'outil Link&Pred

Dans le cas de nos données, les deux algorithmes qui ont montré des résultats efficaces (avec une justesse située entre 92% et 96%) sont : SimRank¹⁴ et CommonNeighbors¹⁵, ils ont donc été implémentés dans l'outil. Les nœuds colorés représentés sur le graphe ci-dessous représentent des liens entre deux intellectuels tandis qu'à gauche, c'est le résultat du calcul appliqué qui apparaît. Ce résultat, comme le montre l'image ci-dessous (voir flèche), sont retournés par l'outil à travers une phrase qui explique le résultat du calcul et propose au chercheur d'explorer une relation entre deux auteurs. Plus le score est supérieur à 1.0 plus il existe de traits communs entre deux intellectuels : le score suggère donc que les deux intellectuels cités partagent des traits communs.



¹³ La présentation de cet outil et de la méthode a été présentée dans le rapport de projet de fin d'études intitulé « Vers un outil intelligent d'exploitation et d'analyse de données d'archives » pour l'obtention du diplôme de « Master en Informatique Option : Génie Logiciel » de Abdelghani LAIFA (Université de Constantine, Algérie / Laboratoire d'Informatique et d'Automatique (LIAS) pour les Systèmes de l'ISAE-ENSMA, Poitiers).

¹⁴ <https://en.wikipedia.org/wiki/SimRank>

¹⁵ <https://neo4j.com/docs/graph-algorithms/current/labs-algorithms/common-neighbors/>

Le résultat de ce premier stage a permis de tester et de confirmer l'utilité des méthodes de prédiction de liens appliquées aux *médiauteurs culturels*.

b) Evolutions de « Link&Pred »

En 2020, avec le soutien du projet e-C@NA¹⁶, un second stage a été mis en place pour poursuivre le travail initié sur l'outil « Link&Pred »¹⁷ et réfléchir aux moyens de faire évoluer l'outil afin qu'il réalise des prédictions plus pertinentes à partir d'un nombre d'attributs (étiquettes des métadonnées) plus important. L'autre objectif du stage était de développer une interface plus « user-friendly » pour l'outil afin de faciliter son exploitation. Il s'agit d'une application web qui comporte déjà un nombre important de lignes de codes.

Lors de ce stage, c'est l'algorithme CNGF¹⁸ qui a été jugé plus performant que Common Neighbors et son implémentation a été testée. Les résultats se sont révélés convaincants mais les expériences doivent être poursuivies en 2021 et 2022 pour parvenir à une proposition fiable.



Interface de prédiction de liens et affichage de graphe

¹⁶ <https://ecana.hypotheses.org/>

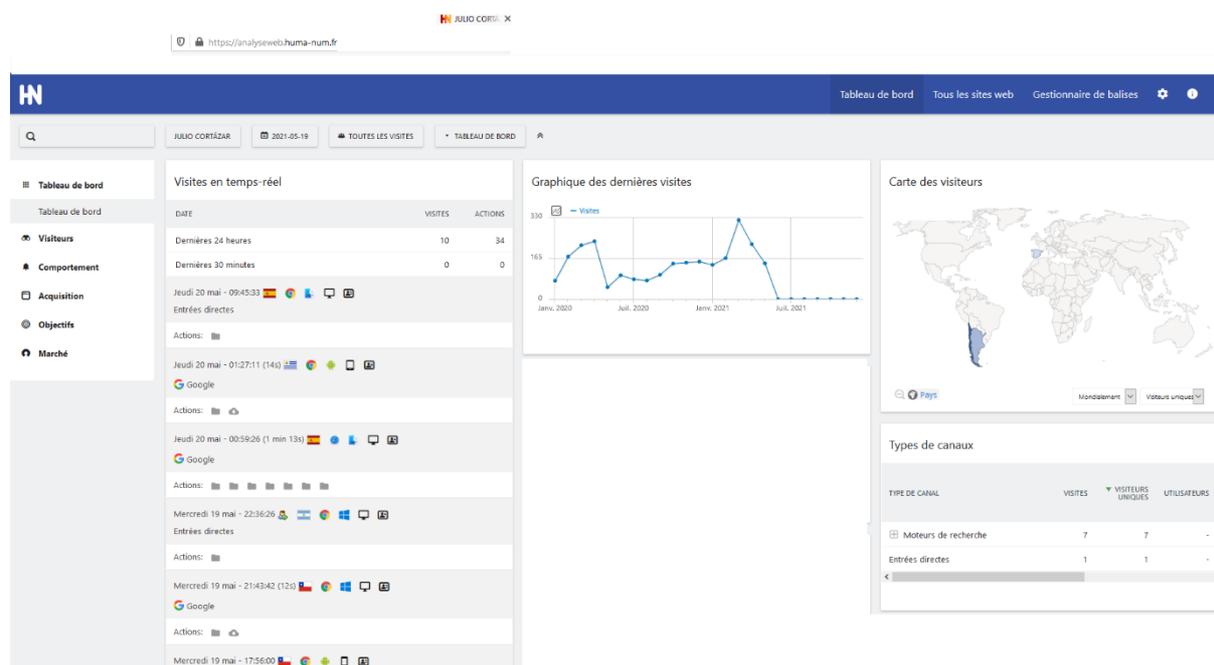
¹⁷ Les partenaires impliqués étaient l'Université de Tizi-Ouzou, Algérie et le Laboratoire d'Informatique et d'Automatique (LIAS) pour les Systèmes de l'ISAE-ENSMA, Poitiers. L'étudiant recruté, Nadjib BENAMROUCHE a travaillé sur la « Prédiction avancée et multidimensionnelle dans le contexte de données d'archives de faible qualité ».

¹⁸ <https://www.hindawi.com/journals/mpe/2013/125123/>

3.4. Conclusions : comment dégager des types ?

Au terme des deux années, le projet « Archivos 3.0 » a résolu un premier problème : celui de l'accès à des documents importants pour la recherche scientifique sur les auteurs latino-américains. Grâce à la numérisation, à la reprise des données et à leur diffusion, les collections ont pu être complétées, mises à jour et les données vont être très prochainement publiées en ligne.

Dans le cas des fonds accessibles en ligne, quelques statistiques de fréquentation fournies par l'outil Matomo (service Huma-Num), permettent déjà de témoigner de l'intérêt de la mise à disposition de ces ressources, notamment pour certains partenaires de l'équipe qui peuvent se connecter depuis leurs pays (Argentine, Chili et Espagne par exemple) et accéder aux données.



Afin d'apporter à la recherche scientifique de nouvelles méthodes, notamment en vue des études sur les *médiateurs des transferts culturels*, le projet a expérimenté une nouvelle proposition théorique, celle de constituer une typologie de ces médiateurs à partir des métadonnées produites sur les fonds d'archives des écrivains. La création de représentations graphiques et l'utilisation d'algorithmes ont permis d'engager le travail d'identification de traits communs qui permettra de soutenir cette proposition typologique tandis que l'utilisation d'algorithmes prédictifs s'est révélée pertinente pour le cas d'études. Enfin, il est apparu nécessaire d'accompagner l'utilisation de ces algorithmes d'un outil sous forme d'application web afin de manipuler les données et de créer des graphes. Ce travail étant encore exploratoire, ni l'outil ni le code informatique écrit ne peuvent encore être partagés.

Durant les prochains mois, dans le cadre du projet « Archivos 3.0 » 2020-2022, le travail se poursuivra sur les archives de l'équipe CRLA-Archivos et nous espérons finaliser un prototype de sorte qu'il puisse être testé avec le public

4. Calendrier du projet 2019-2021

Projet - Archivos 3.0									
Tâche 0 à 3 Livrabale / Personne impliquée	Calendrier	ANNEE 1				ANNEE 2			
		Trimestre 1	Trimestre 2	Trimestre 3	Trimestre 4	Trimestre 5	Trimestre 6	Trimestre 7	Trimestre 8
T 0 : Management, coordination									
Direction du programme / Responsable du Projet									
T 1 : Collecte des données, numérisations									
Plan de classement, du Plan de gestion des données / Responsable du projet + Archiviste CDD		J.Cortazar F.Ainsa	J.Cortazar F.Ainsa	J.Cortazar F.Ainsa	F.Ainsa J.Gilard	F.Ainsa J.Gilard	F.Ainsa J.Gilard	F.Ainsa J.Gilard	F.Ainsa J.Gilard
Saisie des métadonnées / Archiviste CDD		J.Cortazar F.Ainsa	J.Cortazar F.Ainsa	J.Cortazar F.Ainsa	F.Ainsa J.Gilard	F.Ainsa J.Gilard	F.Ainsa J.Gilard	F.Ainsa J.Gilard	F.Ainsa J.Gilard
Numérisations / Stagiaires + Archiviste CDD									
Traductions des métadonnées / Stagiaires									
Contrôle qualité / Responsable du projet + Archiviste				F.Ainsa	F.Ainsa	F.Ainsa	F.Ainsa	F.Ainsa	F.Ainsa
Préservation physique des documents / Responsable du projet + Archiviste + Personnels MSHS									
T2 : Stockage, préservation des données, diffusion des données									
Préparation des paquets pour Nakala + Envoie des données pour Nakalona / Responsable du projet + Archiviste + Personnels MSHS + Prestation informatique							J.Cortazar F.Ainsa J.Gilard	J.Cortazar F.Ainsa J.Gilard	J.Cortazar F.Ainsa J.Gilard
Contrôle qualité / Responsable du projet + Archiviste									
Communication / Stagiaires									
T3 : Analyses et valorisation									
Bilan du ColleX / Responsable du projet									
Suivi de la valorisation / Responsable du projet									
Communication scientifique (publications, événements, etc.) / Equipe scientifique du CRLA-Archivos									

5. Annexes

- 5.1.1. Plan de classement du Fonds Julio Cortázar
- 5.1.2. Plan de classement et inventaire du Fonds Fernando Aínsa (à la date du 12-05-2021)
- 5.1.3. Plan de classement et inventaire du Fonds Jacques Gilard (à la date du 12-05-2021)