Ès lettres

Projet lauréat de l'AAP CollEx-Persée 2019-20, bilan scientifique

Table des matières

1 – Objectifs et organisation 3
1.1 – Contexte et objectifs du projet 3
1.2. - Organisation 3
2 – Signalement et numérisation des thèses 4
2.1 – Catalogage 4
2.2 - Bilan numérisation thèses 4
3 – Base de données 5
3.1 – Cadre technique : Heurist 5
3.2 – Extraction et structuration des données 6
3.3 – Le site en ligne et ses fonctionnalités 7
4 - Valorisation du projet 8
4.1 - Participations à des séminaires ou colloques 8
4.2 - Carnet hypothèses 10
4.3 - Journée d'études, colloque et publications 10
4.4 - Exposition virtuelle 11
Conclusion et perspectives 12

1 – Objectifs et organisation

1.1 – Contexte et objectifs du projet

L'objectif du projet ès lettres est de valoriser le fonds le plus complet de thèses de doctorat ès lettres en France, celui de la Bibliothèque interuniversitaire de la Sorbonne (BIS), en en faisant le point d'entrée pour l'étude de l'interface entre le système de production des savoirs scientifiques et le système de reproduction des élites intellectuelles, en particulier un observatoire de l'institutionnalisation des disciplines. De fait, le doctorat et les thèses restent à ce jour très peu explorés par l'historiographie, par manque d'outils pour aborder une documentation à la fois massive et dispersée – et aussi parce que la mémoire universitaire a eu tendance à dévaloriser les doctorats les plus anciens, sans nécessairement prendre la peine de les lire.

Le projet se concentre sur les thèses soutenues au XIXe siècle. 2 154 thèses soutenues par 1069 individus ont été recensées, d'autres sont encore susceptibles d'être retrouvées. Il s'agit tout d'abord de rassembler des informations relatives aux thèses, actuellement dispersées entre de multiples sources, et de procéder à la numérisation de ces thèses tout en préparant celle de documents qui leur sont associés (notamment les rapports de soutenance). Il s'agit en outre d'élaborer une bibliographie à la fois générale et spécialisée sur ces thèses, de constituer une base de données en ligne et reliée à des référentiels à partir de l'ensemble de ces éléments, et enfin de valoriser scientifiquement ces matériaux.

1.2. - Organisation

Le projet est co-porté par une conservatrice du Service de la valorisation numérique des collections (SERVAL) de la Bibliothèque interuniversitaire de la Sorbonne et un maître de conférences du Centre Lucien Febvre de l'Université de Franche-Comté.

Le dépouillement des sources a été réalisé d'une part par deux ingénieures d'études qui ont successivement travaillé sur le projet et par plusieurs vacataires, financé·e·s par des crédits CollEx-Persée. Les dépouillements aux Archives nationales, principal gisement pour ce projet et institution partenaire dès le début, ont été grandement facilités par les conservatrices des services concernés. La modélisation et l'insertion des données dans la base ont été assurées par les deux ingénieures d'études recrutées, ainsi que par une autre ingénieure d'études en poste à la BIS jusqu'à août 2022. La mise en ligne de la base sur un site dédié a été prise en charge par l'informaticien ingénieur de recherche de la BIS.

Les lots de numérisation ont été préparés, acheminés et contrôlés par la bibliothécaire adjointe spécialisée chargée de la bibliothèque numérique de la BIS au SERVAL, avec le concours de magasiniers et de l'atelier de restauration de la BIS. La numérisation a été effectuée comme prévu sur la filière partenaires de la Bibliothèque nationale de France (BnF), laquelle s'était engagée à numériser 100 000 pages dans le cadre du projet ; il faut souligner que la BnF est par ailleurs un partenaire régulier de la BIS, qui en est « pôle associé ».

Un comité scientifique, composé de vingt chercheur·se·s (universités de Franche-Comté, Strasbourg, Tours, Versailles Saint-Quentin-en-Yvelines, Lille, Le Mans, ENS de Lyon et Ulm, EHESS), et deux conservatrices (Archives nationales et École nationale des chartes) et présidé par le maître de conférences porteur du projet, a été sollicité tout au long du projet pour valider ses orientations, en particulier le modèle de données de la base, et organiser certaines actions de valorisation. Un comité de valorisation s'est également réuni pour prévoir des communications à des événements ou des billets en ligne. C'est ce comité qui a notamment été sollicité pour alimenter le carnet Hypothèses créé pour suivre et valoriser les résultats du projet.

La collaboration avec l'École nationale des chartes (qui mène un projet sur les thèses soutenues à l'École des chartes, Thenc@) et les acteurs.rices du projet Synthèses à Strasbourg a été essentielle pour la réflexion sur la base de données.

2 – Signalement et numérisation des thèses

2.1 – Catalogage

Le catalogage des thèses est un sujet qui intéresse plusieurs autres acteurs de l'information scientifique et technique, au premier rang desquels l'Agence bibliographique de l'enseignement supérieur (Abes). L'Abes a été sollicitée en amont puis en cours de projet pour soumettre les difficultés de catalogage rencontrées. Des échanges fructueux ont permis de lever des ambiguïtés et d'enrichir les données du catalogue collectif national de l'enseignement supérieur, le Sudoc. Les aléas survenus en cours de projet n'ont en revanche pas permis de recontacter les trois SCD (Toulouse Jean Jaurès, Franche-Comté et Aix-Marseille) qui devaient être sollicités pour faire des recherches et compléter éventuellement le corpus par des thèses soutenues dans leur ville. Ils le seront dans la suite du projet, avec les autres SCD repérés comme utiles car conservant des thèses relevant de la période concernée et non conservées à la BIS.

2.2 - Bilan numérisation thèses

Thèse d'Henri Bergson
https://gallica.bnf.fr/ark:/12148/bpt6k1520945m

Le projet prévoyait le début de la numérisation du vaste corpus concerné, soit le traitement de 100 000 vues. D'une part, la numérisation a pu démarrer en amont du démarrage officiel du projet CollEx-Persée, d'autre part, les quantités envoyées ont pu être supérieures aux prévisions. L'objectif est donc largement dépassé puisque 1 062 thèses soit 174 207 vues ont été traitées et 171 thèses soit environ 28 400 vues vont l'être à partir d'août-septembre 2023. Plus de la moitié du corpus est donc traitée.

Les documents sont visibles sur Gallica, océrisés et requêtables par un code

projet unique (« BisThLet »).

3 – Base de données

3.1 – Cadre technique : Heurist

Un outil, Heurist, avait été repéré en amont du projet en vue de créer la base de données du projet. Le choix a été confirmé, l'outil correspondant parfaitement au besoin. Le modèle de données a été discuté en comité scientifique et validé en février 2021, ne subissant que peu d'ajustements par la suite.

Modèle de données du projet ès lettres

3.2 – Extraction et structuration des données

Les données issues des thèses elles-mêmes et des documents trouvés dans les Archives nationales et départementales ont été insérées au fil de l'eau dans la base de données. En priorité, ont été exploitées un certain nombre de cotes relevées aux Archives nationales (dans les séries F/17 et AJ/16 en particulier). Puis l'ingénieure d'études chargée de cet aspect du projet s'est rendue aux Archives départementales de Rennes, Marseille, Caen et Clermont-Ferrand. Six vacataires étudiant·e·s domicilié·e·s sur place ont pour finir été sollicités pour dépouiller les dossiers repérés dans les Archives départementales de Bordeaux, Lille, Lyon et Montpellier. Le porteur scientifique du projet s'est rendu à Dijon, Besançon et Nancy pour compléter le tout.

Le repérage initial était pertinent, c'est-à-dire que les documents les plus utiles se trouvaient bien aux Archives nationales, dans une moindre mesure à Rennes, Marseille et Caen et plus rarement dans les quatre derniers centres d'archives visités. Étaient particulièrement recherchés les rapports de soutenance et dossiers de carrière, et de fait, 939 rapports de soutenance concernant 598 soutenances ont été retrouvés (il y a fréquemment deux rapports, l'un du doyen, à partir de 1839-40, l'autre assez systématique mais officieux, d'un inspecteur général, à partir du courant des années 1860). Mais d'autres documents ont apporté d'utiles informations : correspondances, informations sur le cadre de soutenance (costumes, décor, …), précisions sur la composition du jury et la circulation administrative des thèses, liste de sommes perçues notamment.

Heurist – capture d'écran, exemple d'exploration

Toutes les personnes effectuant les dépouillements remplissaient un même tableau de dépouillement, élaboré en amont. Étaient relevés (si disponibles) : nom, prénom(s), date et lieu de naissance, date de délivrance du diplôme, date(s) de soutenance(s), cote du dossier d'origine dans le fonds d'archive consulté, nature du ou des document(s) retrouvé(s), nom du doyen, nom du secrétaire, nom des membres du jury, présence ou absence de signature, mention obtenue, dispense éventuelle de licence. C'est à partir de ce tableau qu'étaient faites périodiquement par l'ingénieure d'études des extractions pour versement dans la base heurist.

Heurist – capture d'écran, structure des données

Les données dans Heurist ont été structurées autour de six entités, agents (personnes, administration, …), productions (thèses, rapports de soutenances, …), événements (naissance, décès, …), lieux, archives et documentation (c'est-à-dire bibliographie). Elles sont reliées entre elles par des liens, hiérarchiques ou non ; des vocabulaires contrôlés ont été utilisés pour chaque type d'information.

3.3 – Le site en ligne et ses fonctionnalités

La base de données est hébergée par l'université Paris 1 Panthéon Sorbonne, sur un serveur qui a été financé par BIS. La mise en ligne a été assurée par l'informaticien ingénieur de recherche du SERVAL à la BIS. Le choix a été fait d'utiliser l'outil Heurist pour la mise en ligne également, et non d'avoir recours à un outil spécifique, Heurist suffisant aux besoins d'exposition et d'interrogation des données souhaités. Accessible depuis septembre 2022, le site reçoit une fréquentation encourageante, soit environ 15 visites par semaine.

Base de données en ligne du projet ès lettres
https://eslettres.bis-sorbonne.fr/

Il permet un accès aux données par facultés, par thèses, par doctorants et par membres du jury, des recherches dans l'intégralité de la base ainsi que des exports en json et en xml. Chaque fiche générée à la volée après une interrogation permet d'afficher toutes les informations liées. Par exemple, une interrogation sur un doctorant permet d'afficher son portrait, ses informations biographiques, le lien vers toutes les bases bibliographiques ayant une entrée le concernant (idref, persee, …), sa ou ses thèse(s) ainsi que les exemplaires répertoriés dans le Sudoc et les éventuelles numérisations en ligne sur Gallica, sa soutenance de thèse, les membres de son jury, et les éventuels rapports de soutenance.

4 - Valorisation du projet

Les actions de valorisation détaillées ci-dessous ont fait l'objet de publications sur les réseaux sociaux de la BIS, twitter (https://twitter.com/BIS_sorbonne), instagram (https://www.instagram.com/bis_sorbonne/), facebook (http://fr-fr.facebook.com/pages/Paris-France/Bibliotheque-interuniversitaire-de-la-Sorbonne/302275795753), LinkedIn (https://www.linkedin.com/company/biblioth%C3%A8que-de-la-sorbonne/), relayées par les institutions partenaires, ainsi que de publications sur les comptes personnels des protagonistes du projet.

4.1 - Participations à des séminaires ou colloques

Les porteur·se·s de projet ont été très sollicité·e·s pour communiquer sur le projet dans des séminaires ou journées d'études. En effet, nombreux·ses sont les chercheur·se·s et conservateur·rice·s qui travaillent sur le sujet ou ont l'intention de le faire. L'intérêt a porté à la fois sur la collaboration réussie entre recherche et documentation, qui est, de fait, un objectif du GIS CollEx-Persée, sur l'organisation du projet, sur le modèle de données qui porte la base de données mise en ligne et sur les résultats obtenus. Les porteur·se·s de projet

sont intervenus, en sus des manifestations prévues dans le projet (*cf.* 4.3), à neuf reprises :

- 11/12/2020 : communication de Pierre Verschueren « Premiers jalons pour une socio-histoire des docteurs ès lettres : le projet ès lettres », au séminaire d'histoire des sciences humaines et sociales du Centre Alexandre Koyré (EHESS).
- 17/03/2021 : communication de Cécile Obligi « Valorisation numérique, humanités numériques : un cas pratique, le « SERVAL » à la Bibliothèque interuniversitaire de la Sorbonne » et Pierre Verschueren, « Le projet ès lettres », à une table ronde intitulée « Les SHS à l'ère du numérique. Le numérique : facteur d'innovation pour les sciences humaines ? », organisée par des étudiant·e·s de master 2 en Ingénierie, Édition Numérique et Patrimoines de l'Antiquité à nos Jours de l'université de Franche-Comté (captation vidéo en ligne : https://www.youtube.com/watch?v=meQgbuRc2xc).
- 11/05/2021 : communication de Cécile Obligi, Alyx Taounza-Jeminet et Pierre Verschueren, « Projet ès lettres : pour une valorisation numérique des thèses de doctorat ès lettres en France au XIXe siècle » à l'occasion de l'atelier organisé par l'École nationale des chartes dans le cadre de son projet Thenc@ (numérisation des thèses des élèves de l'École des chartes).
- 09/06/2021 : communication en visio de Pierre Verschueren et Cécile Obligi, intitulée « Pour une socio-histoire des docteurs ès lettres : le projet ès lettres », pour le groupe de travail sur l'histoire et la sociologie de l'enseignement supérieur (École normale supérieure de Lyon).
- 26/11/2021 : intervention d'Arnaud Desvignes intitulée « Le "quarteron" des chartistes docteurs ès lettres entre 1815 et 1940 : essai de biographie collective » à la journée d'études « Ouvrir le patrimoine académique : les thèses d'École des chartes accessibles en ligne" organisée par les membres du projet ThENC@.
- 20/06/2022 : communication de Lucie Lachenal-Taballet, Alyx Taounza-Jeminet et Pierre Verschueren « Dépouillement et (re)modélisation : collaborer autour de sources incertaines dans le cadre du projet ès lettres » lors du colloque annuel de la #dhnord « Travailler en humanités numériques : collaborations, complémentarités et tensions ».
- 27 et 28/06/2022 : communication de Pierre Verschueren, « Le projet ès lettres : vers une socio-histoire des docteurs ès lettres » lors du congrès intitulé « L'enseignement supérieur du XIXe siècle à nos jours (France et international) : établissements, acteurs, disciplines, pédagogies » organisé par l''Association transdisciplinaire pour les Recherches Historiques sur l'Éducation (ATRHE) à Lyon (programme : https://enseignemt-sup.sciencesconf.org/data/pages/pgm_long_V9.pdf).
- 11/06/2023 : communication de Pierre Verschueren intitulée « Deux ou trois choses que je sais de lui… le doctorat », lors de la journée nationale du doctorat organisée à Angers par le ministère de l'Enseignement supérieur, de la recherche et de l'innovation.

- 23/06/2023 : communication de Laurie Aoustet et Pierre Verschueren intitulée « Les thèses de doctorat au cœur de la collaboration entre chercheurs et bibliothèques » dans le cadre du séminaire d'histoire des mathématiques de l'Institut Henri Poincaré (https://www.ihp.fr/fr/agenda/autour-des-archives-dans-les-bibliotheques-de-mathematiques).

Une nouvelle intervention, par Laurie Aoustet et Pierre Verschueren, intitulée « Le projet ès lettres : les humanités numériques au service de l'étude du doctorat au XIXe siècle », est prévue le 4 octobre 2023 à l'université Paris-Est Créteil, dans le cadre du colloque « Élaboration d'outils pour l'histoire du livre, de l'écrit et de la lecture à l'ère du numérique : Omeka vs Heurist, xml-TEI et TXM.

4.2 - Carnet hypothèses

Un carnet hypothèses, permettant de suivre la progression du projet, de communiquer dessus et de présenter les premiers résultats obtenus, a été mis en ligne en février 2022 : https://eslettres.hypotheses.org/. Il a permis aux chercheur·se·s membres du comité de valorisation de publier des billets, au porteur·se·s du projet de communiquer sur les actions de valorisation et de faire le lien avec la base de données dès qu'elle a été mise en ligne. En 2022, le carnet a reçu 8 643 visites de 4 361 visiteur·se·s différent·e·s.

Carnet Hypothèses du projet ès lettres
https://eslettres.hypotheses.org/

4.3 - Journée d'études, colloque et publications

Les porteur·se·s de projet ont été sollicité·e·s pour une publication dans la *Revue d'histoire des sciences humaines*. Un article intitulé « Le projet ès lettres : premiers jalons pour une socio-histoire des docteurs ès lettres » rédigé par Pierre Verschueren et Cécile Obligi, est paru en 2021 dans le numéro 39 (https://doi.org/10.4000/rhsh.6672). Écrit au début du projet, l'article expose le contexte, les partenaires, la problématique et les objectifs visés, les méthodes de travail adoptées, ainsi que les premières réalisations.

Le projet a donné lieu comme prévu à une journée d'études internationale intitulée « Histoires de thèses : autour du projet ès lettres » à Besançon le 23/05/2022. Le programme (https://eslettres.hypotheses.org/589) a permis à la fois de présenter les contours, enjeux et premières réalisations du projet et à sept chercheur·se·s de présenter des études de cas sur un docteur, une région, une période, un petit corpus et d'amorcer des réflexions plus transversales sur le sujet.

Le grand colloque conclusif (https://eslettres.hypotheses.org/788) sur appel à contributions évaluées par le comité scientifique), qui a eu lieu à la Sorbonne les 08 et 09/09/2022, a donné lieu à 28 communications de conservatrices ou chercheur·se·s européen·ne·s permettant de commencer à poser les jalons d'une étude sociohistorique des différentes formes de doctorat, s'intéressant aux mécanismes sociaux de formation à la recherche et aux fonctions sociales de ces mécanismes, aux enjeux sociaux et politiques qu'ils constituent, aux parcours et trajectoires des docteurs. Il aborde des questions touchant la construction des disciplines et la mise en place des systèmes universitaires modernes comme leurs métamorphoses les plus récentes, à partir de cas à échelle européenne.

Les riches contributions produites à l'occasion de ces deux événements ont permis de construire un ouvrage qui a été proposé pour une coédition aux Presses universitaires de Franche-Comté et aux Éditions de la Sorbonne. Le manuscrit a été, après expertise, accepté par les deux comités de lecture et sera publié en 2024.

4.4 - Exposition virtuelle

Les premiers résultats scientifiques, en partie exploités sur le carnet hypothèses et dans la publication détaillée en 4.3, sont également présentés dans une exposition virtuelle en ligne sur la bibliothèque numérique de la BIS : https://nubis.univ-paris1.fr/s/theses-doctorats-es-lettres-19-siecle-exposition-devenir-savant/

DEVENIR SAVANTS :
THÈSES ET DOCTORATS ÈS LETTRES AU XIXe SIÈCLE

Introduction | Les cadres institutionnels du doctorat ès lettres au XIXe siècle | L'installation des thèses dans les rayons des bibliothèques universitaires | Une histoire matérielle des thèses de doctorat | Docteurs ès lettres au XIXe siècle : quelques portraits | La fabrique du docteur ès lettres : sur les pas de Camille Jullian | À propos de l'exposition | Docteurs ès lettres au XIXe siècle : quelques portraits

Exposition virtuelle du projet ès lettres – page d'accueil
https://nubis.univ-paris1.fr/s/theses-doctorats-es-lettres-19-siecle-exposition-devenir-savant/

Cette exposition est l'occasion de proposer une histoire matérielle des thèses de doctorat, prenant en compte les dimensions institutionnelles et bibliothéconomiques, articulée à une série de portraits de docteurs représentatifs des évolutions de la thèse au cours du siècle.

Conclusion et perspectives

Le projet ès lettres, dans sa partie co-financée par CollEx-Persée, a atteint ses objectifs de numérisation et de valorisation du corpus : les données sur les thèses, leurs auteurs, leurs jurys de soutenance, leurs conditions de rédaction, soutenance et impression ont été considérablement enrichies, la base de données les regroupant est fonctionnelle, simple d'usage et riche de possibilités d'interrogation, la numérisation a avancé au-delà des objectifs fixés, et les actions de valorisation ont été nombreuses et fructueuses.

Les porteurs souhaitent poursuivre le travail au-delà du projet tel qu'il a été présenté pour financement à CollEx-Persée. En premier lieu, il importe de terminer la numérisation des thèses elles-mêmes (sachant que certaines n'ont pas encore été retrouvées), car le corpus numérisé, même en comptant les documents numérisés par d'autres institutions (21 repérés sur Numelyo par exemple), n'est pas encore complet. Ce travail est déjà engagé, une nouvelle demande de quota sur la filière partenaire de la BnF a été faite et acceptée ; deux nouveaux lots, totalisant près de 30 000 vues, vont partir en numérisation à la fin de l'été 2023. La BIS poursuivra également la numérisation sur ses ressources propres (un marché de numérisation est en cours de montage fin 2023. En outre, une opération de reconnaissance optique de caractères (OCR) pour les documents n'en disposant pas, et de vérification de la qualité de celles qui existent, sera à réaliser. Le corpus ainsi traité se prêtera alors à des analyses lexicales, de la reconnaissance d'entités nommées et du *topic modeling* (permettant de classer automatiquement les thèses par thème).

Il serait aussi envisageable, dans la mesure où la collection de la bibliothèque interuniversitaire de la Sorbonne est, sinon exhaustive, du moins particulièrement riche jusqu'à 1939, de compléter le corpus avec des données sur les thèses du début du XXe siècle, en particulier pour la partie base de données, la partie numérisation posant des problèmes de droits. Enfin, les porteur·se·s réfléchissent à une collaboration franco-allemande, appuyée sur le réseau FID (Fachinformationsdienste für die Wissenschaft), qui ouvrirait la voie à d'intéressantes études transfrontalières – des dispositifs de circulations des thèses entre monde francophone et espace germanique fonctionnant dès 1885.