

Appel à candidatures

Cartographie des thèses françaises numérisées

Objet

Développement d'une vision globale chronologique et disciplinaire de l'état de numérisation rétrospective des thèses soutenues entre 1806 et 2006 dans les établissements d'enseignement supérieur et de recherche français

Établissement

Service Commun de Documentation,

Université de Lille

Rue du Barreau – CS 90099 - 59652 Villeneuve d'Ascq Cedex

Date limite de réponse :

30 juin 2026 à 12h

Pour obtenir toute information complémentaire, les opérateurs économiques ont la possibilité de s'adresser à :

Karolina SUCHECKA, chargée de mission Numérisation enrichie de CollEx-Persée

Service Commun de Documentation, Université de Lille

Rue du Barreau – CS 90099 - 59652 Villeneuve d'Ascq Cedex

karolina.sucheka@univ-lille.fr / 03 62 26 97 55

Contexte de la mission

CollEx-Persée¹ est une infrastructure nationale de recherche en information scientifique et technique, organisée sous la forme d'un groupement d'intérêt scientifique. Elle a pour mission de développer, mutualiser et coordonner les services documentaires au service de la recherche.

Le programme de Numérisation enrichie, co-porté par l'Université de Lille et Persée, est l'un des quatre programmes CollEx-Persée. Il a notamment pour objectif de développer, de manière concertée, l'offre de littérature scientifique numérisée. Pour ce faire, il est nécessaire, dans un premier temps, de disposer d'un état des lieux des collections déjà numérisées afin d'analyser leur couverture, de faire le lien avec les besoins disciplinaires et d'identifier les collections à numériser en priorité.

Les thèses de doctorat, ainsi que les thèses anciennes, représentent un objectif important pour la numérisation de la littérature scientifique. L'élaboration d'une stratégie de numérisation nécessite donc un travail préalable de recensement afin de disposer d'une vision globale de l'existant, encore incomplète aujourd'hui.

En effet, les études consacrées à la numérisation des thèses se sont jusqu'alors concentrées sur les projets portés par les établissements d'enseignement supérieur et de recherche (ESR), et sur les thèses soutenues à partir de 1985.

Ainsi, en janvier 2020, l'Agence bibliographique de l'enseignement supérieur (Abes) a conduit une enquête auprès de 23 établissements de l'ESR, visant à établir une typologie des fonds de thèses numérisées et à dresser un état des lieux des modalités techniques et organisationnelles mises en œuvre².

Cette enquête a constitué un point de départ pour une étude de faisabilité d'un projet national de numérisation rétrospective des thèses de doctorat, commandée en 2021 par l'Université de Lille et l'Abes, et réalisée par un groupe d'élèves conservateurs de l'École Nationale Supérieure des Sciences de l'Information et des Bibliothèques (ENSSIB). En s'appuyant notamment sur les fonds de thèses reproduites sur microfiche par l'Atelier National de la Reproduction des Thèses (ANRT) et conservés à l'Université de Lille (environ 230 000 références), cette étude a permis d'évaluer les conditions de mise en œuvre d'un tel projet, en envisageant trois scénarios : (1) l'ouverture de l'accès à la totalité de la collection de l'ANRT, (2) une numérisation disciplinaire et

¹ Voir <https://www.collexpersee.eu/>.

² Voir « La numérisation des thèses de doctorat françaises – État des lieux. Enquête 2020 », L'Abes, 2020. URL : <https://abes.fr/wp-content/uploads/2020/10/la-numerisation-des-theses-de-doctorat-francaises-etat-des-lieux-enquete-2020.pdf>.

décentralisée, et (3) le rassemblement des fichiers numériques existants en vue de développer de nouveaux usages. Ce dernier scénario, jugé le plus réaliste, a fait l'objet d'une analyse approfondie.

Dans le prolongement de ces travaux, le groupe de travail CollEx-Persée *Données au service de la numérisation* poursuit actuellement la collecte des données issues de divers projets de numérisation des thèses (thèses numérisées déposées sur TEL/HAL, fonds d'établissements partenaires, projets de recherche portant sur les thèses anciennes, etc.).

Objet de la mission

La mission a pour objectif, en s'appuyant sur les données déjà collectées et sur les conclusions des études précédentes, de produire un état des lieux de la numérisation :

- des thèses de doctorat françaises, en version de soutenance, datant de 1985 à 2006 et déposées dans un format autre que numérique ;
- des thèses anciennes (ès lettres, ès sciences, etc.), thèses de troisième cycle et thèses d'État soutenues en France entre 1806 et 1985.

Les autres types de thèses (d'université, de licence, d'exercice, d'école, et d'habilitation à diriger les recherches, etc.) n'entrent pas dans le périmètre de cette étude.

L'étude prendra en compte les activités de numérisation de thèses menées par l'ensemble des acteurs concernés (bibliothèques universitaires, interuniversitaires ou de recherche, bibliothèques nationales, infrastructures de recherche, plateformes numériques, etc.).

Le programme CollEx-Persée :

- accompagnera la collecte des données en mobilisant ses partenaires ;
- fournira au prestataire, sous la forme d'un tableau Excel, un recensement des thèses (version de soutenance) ainsi que leurs reproductions signalées dans le Sudoc.

Le prestataire :

- complétera le recensement des thèses numérisées :
 - au moyen d'une enquête diffusée auprès des acteurs identifiés ;
 - en moissonnant les données disponibles en ligne (bibliothèques numériques, plateformes de numérisation, catalogues, etc.) ;
- concaténera les données collectées au sein d'une base de données compatible avec les outils de visualisation des données (Lodex, Metabase, etc.).

La mission vise ainsi à établir un panorama aussi complet que possible de l'état de numérisation des thèses, en produisant des données à la fois visualisables et exploitables à des fins d'analyse. À ce titre, elle devra permettre de répondre aux questions suivantes :

- que recouvre l'existant : qu'est-ce qui a été réalisé, qu'est-ce qui est en cours et qu'est-ce qui reste à entreprendre, selon les disciplines ?
- dans quelles perspectives s'inscrivent les projets de numérisation (valorisation, conservation, histoire des universités, histoire des sciences, projets de recherche spécifiques, etc.) ?
- quelles métadonnées sont disponibles, quel est leur niveau d'interopérabilité, et quelles difficultés spécifiques apparaissent lors des opérations de collecte et d'agrégation ?
- quelles ont été les options retenues pour la diffusion des collections de thèses numérisées ?

Les premiers livrables attendus sont :

- un recensement exhaustif des projets de numérisation des thèses permettant de répondre aux questions identifiées ci-dessus ;
- une base de données aussi complète que possible des thèses numérisées.

La structuration des métadonnées permettra notamment le croisement des deux bases.

Dans un second temps, le prestataire conduira une phase d'analyse complémentaire, dont l'orientation sera définie en concertation entre le prestataire et le commanditaire, en fonction des résultats analytiques de la première phase. Celle-ci pourra porter sur :

- une proposition des visualisations permettant d'analyser les corpus déjà numérisés selon différents prismes ;
- une étude des lacunes ;
- une analyse disciplinaire ;
- un focus chronologique, etc.

Descriptif de la mission

Le prestataire sera chargé de mettre en œuvre les tâches suivantes :

- Assurer la collecte, le croisement et l'enrichissement de données relatives aux projets de numérisation des thèses originelles en France, intégrant notamment les informations suivantes pour chaque projet :
 - L'établissement en charge de la numérisation ;
 - La volumétrie ;
 - L'objectif du projet (projet de recherche, conservation de fonds, valorisation, etc.) ;

- Les formats disponibles (image, PDF, texte...)
- La liste des métadonnées renseignées ;
- Les modalités de diffusion de la collection.
- Formuler des analyses critiques sur les enjeux, les limites et les difficultés particulières liés à cette collecte ;
- Concevoir et structurer une base de données des thèses numérisées intégrant notamment les informations suivantes pour chaque thèse :
 - Le titre ;
 - Le nom de l'auteur ;
 - Le nom du directeur ou de la directrice de thèse ;
 - La date de soutenance ;
 - Le Numéro National de Thèse (NNT), si connu ;
 - La langue de la rédaction ;
 - La ou les disciplines ;
 - Les mots-clés et le résumé ;
 - Le type de thèse (thèse d'université, de 3^{ème} cycle, de doctorat, etc.) ;
 - La date de numérisation ;
 - Les modalités d'accès (ouvert ou non) ;
 - Les éléments techniques (numérisation de micro-fiche ou de thèse papier, formats...) ;
 - Le type d'entrepôt hébergeant la thèse numérisée (plateforme universitaire, site Omeka, bibliothèque numérique de type Gallica en marque blanche, etc.) ;
 - Les modalités de stockage et archivage (serveurs internes de l'établissement, archivage pérenne, etc.) ;
 - L'établissement de soutenance ;
- Accompagner CollEx-Persée dans le choix d'outils de visualisation permettant l'exploration des données selon des axes chronologiques et disciplinaires ;
- Conduire une analyse approfondie sur un sous-corpus, défini à partir des résultats globaux observés lors de la première phase de la mission.

Acteurs de la mission

Cette mission s'inscrit dans les travaux du groupe *Données au service de la numérisation*, co-porté par l'Université de Lille et par la Bibliothèque Interuniversitaire de la Sorbonne (BIS). Ce groupe de travail vise à définir des démarches, des méthodes et des outils pour mieux collecter et interpréter les données relatives aux collections numérisées de la littérature scientifique et à leur usage afin d'éclairer les choix en matière de numérisation. Il est composé des professionnels de l'information des universités françaises interdisciplinaires (Université de Lille, Université Grenoble-Alpes, Paris

Nanterre...), des bibliothèques spécialisées (BIS, BIU Cujas) et des opérateurs nationaux (Persée, Inist).

Les interlocutrices principales du prestataire seront Laurie Aoustet, Madeleine Géroudet et Karolina Suchecka (personne contact : karolina.suchecka@univ-lille.fr). Les membres du groupe de travail échangeront avec le prestataire lors des comités de pilotage mensuels.

Calendrier indicatif de la mission

La mission est établie pour la durée suivante : septembre 2026-avril 2027.

Un premier rendu est attendu début 2027 pour décider du déroulement de la seconde phase du projet.

Calendrier prévisionnel de la prestation :

Phase 1 (septembre 2026-janvier 2027)

- CP 1 (septembre 2026): définition de la méthodologie précise, établissement du questionnaire d'enquête ;
- CP 2 (octobre 2026): lancement de l'enquête ;
- CP 3 (janvier 2027): livraison des bases de données (projets de numérisation et thèses numérisées).

Phase 2 (février-avril 2027)

- CP 4 (février 2027): définition du scénario d'approfondissement ;
- CP 5 (avril 2027): livraison de la base consolidée, conclusions et recommandations.

Montant de la mission

Le montant de la mission ne pourra pas dépasser 25 000 € TTC. Toute offre dépassant ce montant sera considérée comme irrecevable.

Composition du dossier de candidature

Le candidat à la réalisation présentera les éléments suivants :

- la présentation du candidat : structure et organisation, équipe mobilisée, etc. ;
- description de l'approche envisagée :
 - identification des enjeux et risques ;
 - méthodologie (stratégie de collecte, modalités d'enquête) ;
 - moyens techniques (outils et traitements envisagés) ;
 - planning et la répartition des tâches ;
 - livrables (description, formats, calendrier) ;
- offre financière (devis détaillé, ventilation par poste) ;
- références des expériences précédentes.

Le dossier de candidature de 20 pages maximum (hors références) est attendu pour le 26 juin à 12h dernier délai. Les résultats seront annoncés à la mi-juillet.

Critères de sélection

Les candidatures seront évaluées selon les critères suivants :

- compréhension du besoin (2,5/25 points) ;
- qualité méthodologique (7,5/25 points) ;
- expérience dans les projets similaires (2,5/25 points) ;
- pertinence des outils proposés (7,5/25 points) ;
- offre financière (5/25 points).

Confidentialité

Le prestataire est tenu à une obligation de confidentialité concernant l'ensemble des informations, données, documents et résultats portés à sa connaissance dans le cadre de la mission. Il ne peut les communiquer à des tiers ni les utiliser à d'autres fins que l'exécution de la prestation sans l'accord préalable du commanditaire. Conformément aux articles 34 à 38 du CCAG-PI approuvé par l'arrêté du 30 mars 2021, les résultats de l'étude et les droits d'utilisation associés sont acquis au commanditaire pour les besoins de ses missions. En application des articles 27 à 30 du même CCAG, le prestataire garantit la conformité des livrables et procède, à ses frais, aux corrections nécessaires en cas d'erreur ou de non-conformité constatée lors de leur vérification. Cette obligation de confidentialité demeure applicable après l'achèvement de la mission.