

SYNTHESE

de l'étude de besoins

relative aux solutions techniques

utilisées par les porteurs de projets

de numérisation de l'ESR



GT Numérisation CollEx-Persée
2020-2021

En 2020-2021, le GT Numérisation a réalisé une étude sur les solutions techniques utilisées par les structures documentaires (BU, BIU, Bibliothèques de Grands Établissements), les MSH et les laboratoires, au cours de leur projet de numérisation, afin d'évaluer la réponse de ces outils aux besoins initiaux, ainsi que leurs limites et leur potentiel de mutualisation à l'échelle nationale.

Cette étude s'inscrit dans le premier axe de la politique nationale de numérisation portée par CollEx-Persée, qui a comme ambition de soutenir la production et la diffusion de corpus numérisés, en garantissant l'interopérabilité et la pérennité de ces corpus. Cet axe s'appuie sur une action de structuration technologique dont les objectifs sont notamment :

- le soutien à la mutualisation, la création et l'évolution d'outils utilisés lors d'un projet de numérisation ;
- la réduction des coûts de développement et de maintenance ;
- le partage de compétences ;
- la convergence et la visibilité de l'ensemble des collections numérisées en garantissant que les projets de numérisation puissent s'adosser les uns aux autres ;
- la promotion des standards de la science ouverte ;
- la recommandation d'outils aux porteurs de projets de numérisation.

Cette étude menée auprès d'une trentaine de porteurs de projets de numérisation de l'Enseignement Supérieur et de la Recherche avait pour double objectif de recenser des besoins n'ayant pas encore trouvé de solutions techniques adéquates et de repérer des outils pouvant être soutenus par CollEx-Persée dans la perspective de construire une chaîne technologique solide, constituée de briques interopérables, mobilisables pour tout ou partie par les structures de l'ESR aguerries ou débutantes en matière de numérisation.

I. Méthodologie de l'étude

L'étude, envisagée dès le départ comme une étude qualitative, visait à repérer des signaux faibles par le biais d'entretiens semi-dirigés réalisés avec les porteurs de projets de numérisation identifiés au sein des réseaux professionnels des membres du GT. Dans un premier temps, une modélisation des actions réalisées lors d'un projet de numérisation a été établie allant de la sélection des documents à leur diffusion, en passant par la numérisation proprement dite, mais également par les enrichissements et la préservation numérique. A partir de ce modèle, une grille de recensement des solutions techniques mobilisées par les établissements interrogés a été élaborée. Cette grille a été envoyée aux personnes interrogées deux semaines avant l'entretien, puis renvoyée remplie aux enquêteurs qui ont ainsi pu préparer l'entretien, en ciblant leurs questions sur les outils présentant le plus d'intérêt en termes de retour d'expérience qualitatif permettant d'envisager une mutualisation ou des développements. Cette méthodologie a d'abord été testée à l'automne 2020 au sein des établissements des membres du GT (dix établissements interrogés) puis après quelques ajustements, elle a été utilisée de janvier à avril 2021 pour réaliser l'enquête auprès d'une vingtaine de structures documentaires relevant de l'ESR. Ainsi, 31 structures ont été interrogées, se répartissant comme suit :

- 16 bibliothèques universitaires et interuniversitaires
- 6 bibliothèques et services documentaires de grands établissements et autres établissements
- 4 Maisons des Sciences de l'Homme
- 5 Unités Mixtes de Recherche et Unités d'Appui à la Recherche

Le GT s'est réuni une fois par mois depuis septembre 2020 pour élaborer la méthodologie de l'enquête et partager et analyser les résultats des entretiens menés.

L'analyse des résultats a eu lieu entre mai et juillet 2021. Elle a consisté à extraire des grilles d'enquête les retours d'expériences sur les solutions techniques mobilisées et les besoins transversaux exprimés, puis à les regrouper en thématiques.

En annexes : *Actions des processus de numérisation, Actualité lancement enquête publiée dans le site CollEx-Persée, Grille d'enquête recensement et entretien.*

II. Constats

L'un des premiers constats fait par le GT Numérisation est la maturité de la numérisation au sein des établissements interrogés. La satisfaction est plutôt bonne pour les outils utilisés, peu de besoins de développements sur les outils eux-mêmes se sont exprimés. Les principaux besoins techniques exprimés portent sur la gestion des processus et des relations avec les prestataires, pour lesquels NumaHOP est envisagé fréquemment comme une réponse adéquate. La majorité des projets d'acquisition ou d'installation de nouvelles solutions portent en effet sur NumaHOP et Omeka S. Des demandes sont faites sur le développement de connecteurs entre NumaHOP et certains outils de diffusion. Par ailleurs, les projets d'acquisition ou d'installation de nouvelles solutions portent également fréquemment sur l'acquisition de logiciels d'OCR et sur l'installation de serveur IIIF.

Ce peu d'attentes sur les solutions techniques de dématérialisation, de post-traitement, de contrôle qualité, de mise en ligne et de diffusion est confirmée par l'[enquête réalisée par la commission SSI de l'ADBU](#) portant sur l'ensemble des applications métiers utilisées en bibliothèques, incluant la numérisation : très peu de réponses ont été collectées sur la partie numérisation alors que près de 70 établissements ont répondu.

Ce sont donc essentiellement des besoins transverses qui se sont exprimés :

- de façon quasi systématique :
 - o des besoins de partage d'expérience, de compétences, de veille technologique. Savoir qui utilise quel outil, comment, quels développements ont été faits, comment mener les discussions avec les prestataires, quelles procédures sont mises en place. Mais également des besoins en formation et en guides de bonnes pratiques sur les langages de description, l'utilisation et l'installation d'outils, l'usage des formats ont été exprimés
 - o des besoins en soutien financier pour la préservation numérique

- de façon régulière
 - o des besoins en expertise juridique pour la collecte des droits de diffusion et d'exploitation des contenus auprès des auteurs et ayant droits, pour les imprimés mais aussi pour les archives de chercheurs ou les films. Près de la moitié des bibliothèques interrogées limitent leurs projets de numérisation aux documents libres de droits pour ne pas avoir à se lancer dans une collecte qu'elles ne savent pas comment réaliser et qui nécessitent beaucoup de temps d'instruction et de suivi. L'autre moitié a trouvé des solutions variées pour réaliser cette collecte. Les refus signifiés par les auteurs et ayants droits les conduisent à envisager des solutions de diffusion intra-muros
 - o des besoins globaux en matière de gestion des archives numériques natives et de numérisation d'objets 3D. Dans ce cas, il s'agit tout à la fois d'avoir des guides de bonnes pratiques et des référentiels pour la numérisation ou la gestion de ces documents mais également d'avoir une chaîne opérationnelle et mutualisable ainsi que des formations et des partages d'expériences. En effet, les établissements interrogés numérisent majoritairement des imprimés (incluant les documents iconographiques) et des manuscrits et ponctuellement des archives, des objets, des documents audios et vidéos, des plaques de verres. La numérisation de ces derniers types de documents est donc moins connue et les compétences moins répandues
 - o des besoins portant sur des outils d'OCR, de HTR, de transcription, de translittération, d'annotation permettant de traiter tous types de documents, d'écritures, de langues
 - o des besoins de tables de correspondances entre différents formats de métadonnées permettant le transfert des données d'un outil à l'autre, sans perte

- de façon ponctuelle
 - o des besoins portant sur la visibilité des projets de numérisation
 - o des besoins portant sur la visibilité des collections numérisées et l'interopérabilité des bibliothèques numériques

L'étude de besoins a été menée dans une logique de repérage de grandes tendances, aussi les constats détaillés ci-dessus mériteraient d'être consolidés et précisés, chacun, par une enquête quantitative.

III. Recommandations d'actions

Plusieurs actions sont envisageables, transverses aux différentes thématiques, d'autres sont plus spécifiques. Toutes mériteraient d'être instruites, budgétées, planifiées :

- Transversales, afin de répondre aux besoins de veille technologique, de partage d'expérience, de nécessité de montée en compétences sur les questions de numérisation 3D, de gestion des archives, d'expertise juridique, d'OCR. L'enjeu est de permettre aux acteurs d'avoir une meilleure connaissance des usages et des développements réalisés par les différents porteurs de projets de numérisation de l'ESR et de favoriser le développement de synergies. Quelques actions possibles :
 - o création d'un réseau des référents numérisation et d'une liste de diffusion
 - o organisation de formations/ateliers/webinaires/journées professionnelles
 - o repérage, recommandation et rédaction de guides de bonnes pratiques
 - o réalisation d'un recensement des solutions techniques mobilisées par les porteurs de projets de numérisation de l'ESR

- Expertise juridique : l'enjeu est de favoriser la numérisation de corpus actuellement sous droits et de permettre ainsi une continuité entre les collections numérisées et les collections numériques natives. Quelques actions possibles :
 - o création d'un groupe d'experts accompagnant les structures documentaires et proposant une assistance juridique
 - o création d'une cellule centralisée gérant la collecte de droits de tous types de documents
 - o financement de l'acquisition des droits auprès des auteurs et ayants droit

- Numérisation d'objets 3D : l'enjeu est d'accompagner les structures documentaires et les unités de recherche dans la numérisation de ce type de documents. Les projets sont fréquemment reportés par manque de repères, de modèles, de référentiels pour orienter l'élaboration des projets de numérisation.
 - o constitution d'une chaîne de traitement dédiée (identification, description, numérisation, diffusion)

- Gestion des archives scientifiques (imprimées, manuscrites, numériques natives, audio, video) : l'enjeu est d'accompagner les structures documentaires dans la gestion de ce type de documents. Les projets sont fréquemment reportés par manque de repères, de modèles, de référentiels pour orienter l'élaboration de ces projets de numérisation/intégration.
 - o constitution d'une chaîne de traitement dédiée (classement, édition, description, numérisation le cas échéant, transcription, enrichissement, exploitation, gestion des droits, diffusion)

- OCR/HTR, transcription, translittération, annotation : c'est une condition indispensable à l'exploitation des contenus, encore très largement prospective au sein des bibliothèques mais déjà investie par les équipes de recherche notamment dans le champ des humanités numériques. L'enjeu est de donner de la visibilité sur les différents outils développés, de permettre leur valorisation et leur mutualisation, de participer à des projets reposant sur l'intelligence artificielle et notamment l'apprentissage par la machine, de permettre l'émergence d'un corpus de données suffisamment homogène pour permettre son exploitation dans des outils variés. Quelques actions possibles :
 - o création d'un mécanisme de soutien financier au développement et à l'acquisition d'outils d'OCR/HTR adaptés aux différents types de documents
 - o conception d'un plan d'OCRisation rétrospective massive au sein des bibliothèques

- Tables de correspondances entre différents formats de métadonnées : cette thématique rejoint l'objectif de permettre l'émergence d'un corpus de données suffisamment homogène pour permettre son exploitation dans des outils variés. L'enjeu est de favoriser l'indépendance des données par rapport à leur contexte de production ou de diffusion, de dissocier les données des outils qui ont servi à les produire, les exploiter, les diffuser afin de permettre leur visibilité et leur interopérabilité. Quelques actions possibles :
 - o réalisation d'un recensement des tables de correspondances existantes et mutualisables
 - o promotion du développement de formats pivots et de procédures partagées de transfert d'un format à l'autre

- Préservation numérique :
 - o mise en oeuvre de l'AMI porté par le GT Préservation
 - o valorisation des logigrammes réalisés par le GT Préservation

IV. Suite donnée

Ces recommandations ont été présentées au GIS CollEx-Persée en juillet 2021.

Le GT Numérisation a inscrit sur sa feuille de route 2021-2022 la mise en place des recommandations liées à la veille technologique et au partage d'expériences et de compétences.

Version mise à jour le 13 septembre 2021

Héloïse Marill et Laurie Aoustet

Pour et Co-pilotes du #GTNumérisation