

NumaHOP, une plateforme de gestion de contenus numérisés

Projet collaboratif développé en open source, NumaHOP offre une chaîne unique, simplifiée et largement automatisée, permettant de gérer toutes les étapes d'un projet de numérisation, de l'import de notices à la diffusion et à l'archivage.

NumaHOP est un projet porté par trois établissements pilotes, bibliothèque Sainte-Geneviève, bibliothèque de Sciences Po-Paris, Bulac, dans le cadre de la Comue Sorbonne Paris Cité, avec le soutien financier du Département de Paris.

Investis dans différents projets de numérisation, ces trois établissements ont rapidement fait le constat de la multiplicité des outils développés localement et méconnus des autres bibliothèques, de l'hétérogénéité des traitements apportés aux différentes étapes de la numérisation suivant les contextes, et de la complexité des différents processus engagés. L'idée de développer une chaîne unique, simplifiée et largement automatisée pour gérer toutes les étapes des projets de numérisation a émergé, après une comparaison des outils existants qui ne répondaient pas pleinement aux besoins des bibliothèques.

Suite à un appel d'offre mené en 2015-2016, le projet de développement a été confié à la société Progilone, en étroite collaboration avec les trois établissements réunis en groupe métier et en comité de pilotage.

UN OUTIL INTÉGRATEUR

L'objectif attendu du projet était triple. En premier lieu, le logiciel développé devait être open source pour être diffusé, utilisé et amélioré le plus largement possible par l'ensemble des établissements intéressés, non seulement dans la sphère universitaire et culturelle (bibliothèques, archives, musées) mais aussi bien au-delà (associations, entreprises). La plateforme de gestion des contenus numérisés devait aussi permettre de gérer, en un point d'entrée unique, l'ensemble de la chaîne de numérisation, en prenant notamment en compte les étapes d'imports en amont, de diffusion en aval, et les interactions avec les éventuels prestataires de numérisation. Compte tenu de la diversité des utilisateurs envisagés, le logiciel devait être le plus « personnalisable » possible, par les agents métiers, pour répondre aux différents enjeux et besoins locaux. Enfin, l'outil devait être accessible le plus simplement possible afin de favoriser son appropriation par l'ensemble des agents, y compris sur des tâches habituellement identifiées comme techniques.

Désormais, l'usage quotidien de NumaHOP par les équipes des établissements pilotes a permis de constater rapidement la simplification des chaînes de numérisation. Cette plateforme, accessible en full web, permet en effet de gérer toutes les étapes de la chaîne, de l'import des notices et du constat d'état des

documents physiques à la diffusion et à l'archivage grâce à un interfaçage largement automatisé avec les différents acteurs impliqués (Abes, prestataires de numérisation, bibliothèques, diffuseurs, Cines).

QUATRE MODULES FONCTIONNELS

NumaHOP est composé de quatre principaux modules fonctionnels. Le premier module constitue l'interface d'import. Il permet notamment de convertir automatiquement des notices au format Unimarc ou EAD¹ dans des formats interoperables : Dublin Core, Dublin Core qualifié. Les mappings sont personnalisables selon les spécificités de chaque établissement. Ce module permet également d'importer les fichiers images, qu'ils soient livrés par un prestataire de numérisation ou directement produits par l'établissement. Plusieurs formats d'image sont pris en charge par NumaHOP : PDF, TIFF, PNG, JPG, JP2000, GIF et SVG. Lors de l'import, des contrôles automatisés sont réalisés par NumaHOP, permettant de détecter d'éventuelles erreurs de formats, de résolution, de compression, de profil de couleur, ou de nommage (séquençage, casse).

Le second module permet la gestion interne des documents, répartis au sein de projets, lots et trains. Une interface dédiée à la rédaction de constats d'état offre la possibilité de consigner les différents états du document, avant et après sa numérisation. Cette interface est paramétrable par établissement et offre la possibilité d'utiliser un vocabulaire uniformisé et contrôlé. Divers outils de gestion de projets (statistiques, tableaux de bord) sont associés à ce module, pour faciliter le recueil d'indicateurs d'activités pertinents.

Le contrôle qualité des images et métadonnées est réalisé au sein d'un troisième module. L'implémentation d'une visionneuse adaptée aux besoins des contrôleurs permet de vérifier sur un même écran la qualité des images produites, les métadonnées qui leur sont associées (table des matières, OCR²), ainsi que divers points de vigilance indiqués dans le constat d'état du document. Lors de la conception de NumaHOP, une attention particulière a été portée à ce module et à ses performances. La création, en temps réel, par le logiciel, de fichiers dérivés dans une résolution adaptée au zoom, permet désormais une très grande rapidité d'affichage. Les erreurs sont signalées directement depuis l'écran de contrôle et permettent l'envoi automatisé au prestataire d'un rapport de contrôle normalisé indiquant les vues à reprendre.



[1] EAD : Encoded Archival Description

[2] OCR : Optical Character Recognition

[3] METS : Metadata Encoding and Transmission Standard

Enfin, le dernier module concerne les fonctionnalités d'export des fichiers images ou métadonnées. NumaHOP permet d'exporter de façon automatisée un document, à la fin du workflow, à la fois vers une plate-forme de stockage local ou d'archivage (Cines), et vers les plates-formes de diffusion (*Internet Archive*, *Omeka*). Les fichiers métadonnées attendus pour les différents types d'exports (par exemple sip.xml et METS³ pour le Cines, OCR pour la diffusion) ainsi que les fichiers images dérivés sont produits automatiquement lors de l'export. Grâce à l'automatisation de ces tâches, les établissements peuvent ainsi disséminer largement et de manière systématique l'ensemble de leurs contenus numérisés.

La mise en place d'un workflow configurable, adapté à chaque projet permet de guider l'utilisateur à travers ces différents modules, selon son profil et ses attributions.

AUTOMATISER, FLUIDIFIER, COLLABORER: ET HOP !

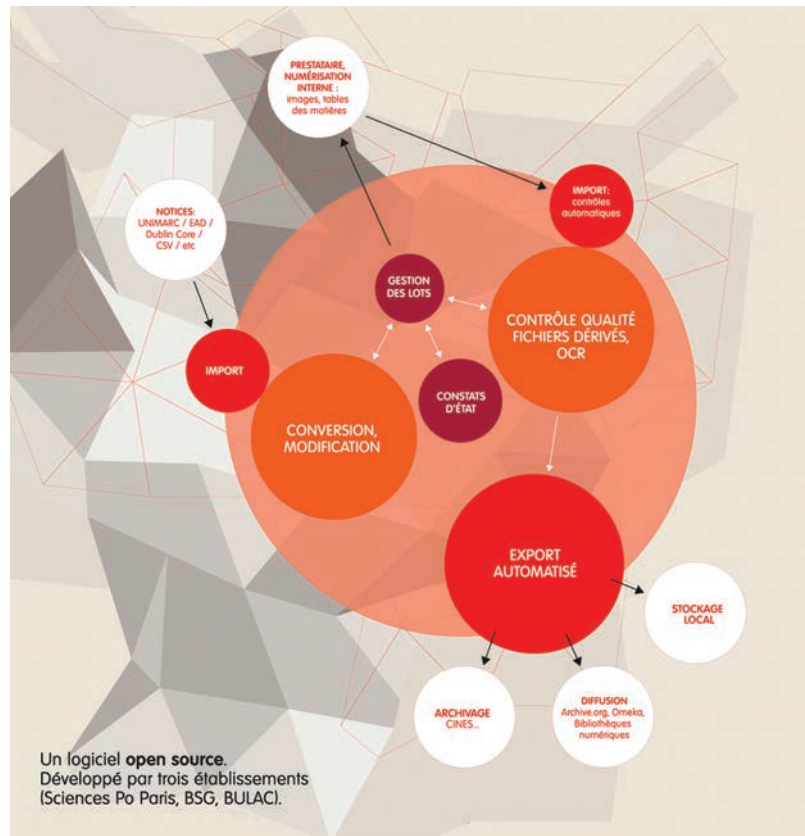
Quel bilan tirer au terme d'un an d'utilisation en production dans les établissements pilotes ? Des gains de temps notables ont été observés, en particulier à deux étapes de la chaîne de numérisation, auparavant très chronophages : la conversion des métadonnées et le contrôle qualité.

La conversion automatique de données bibliographiques permet de se contenter désormais de simples vérifications lors de l'import de notices. Les métadonnées produites dans l'application sont conformes aux normes (Abes, BnF, Cines, etc.), ce qui offre des fichiers d'export directement injectables dans les bibliothèques numériques et au Cines. Le contrôle qualité est aujourd'hui plus fiable et plus rapide, grâce aux contrôles automatiques exhaustifs à l'import et à la performance d'affichage des images. La génération automatique de bordereaux normalisés de contrôle a également permis de fluidifier les procédures de reprise.

Parallèlement, l'utilisation de NumaHOP a permis une simplification des appels d'offres, en limitant considérablement le nombre de tâches attendues des prestataires de numérisation.

En 2020, les établissements pilotes ont constaté une diminution des coûts des prestations de numérisation de 30 à 60 % selon la typologie des documents numérisés. Des prestations variées, auparavant dévolues aux sociétés de numérisation (génération de fichiers dérivés, création de paquets METS, génération et dépôts automatisés des paquets pour l'archivage pérenne au Cines, génération de l'OCR) sont désormais prises en charge directement dans NumaHOP. Au sein des trois établissements, l'utilisation de NumaHOP a permis de recentrer l'activité du prestataire de numérisation sur la prise de vue, tout en reprenant la main sur la gestion et la production des métadonnées.

Ce constat tient notamment au fait que l'application dispose d'une interface ergonomique, qui a permis une acculturation rapide des personnels, y compris sur des tâches auparavant identifiées comme trop techniques.



La mise en production de NumaHOP a permis de ne plus limiter les tâches liées à la numérisation à un petit nombre d'agents aguerris à l'utilisation d'outils techniques parfois rebutants au premier abord. L'intervention de personnes, issues de services distincts, au sein d'un même outil est facilitée par l'utilisation du tableau de bord et des possibilités de « reporting » qui permettent à l'ensemble des intervenants d'avoir une vue globale sur l'avancée d'un projet.

Un des objectifs de NumaHOP réside dans l'utilisation de méthodes de travail standardisées, tout en offrant la souplesse d'adapter l'outil aux réalités locales. L'instance mutualisée qui dessert les trois établissements a ainsi conduit à une harmonisation des pratiques, tout en maintenant des paramétrages distincts pour chacun des modules par établissement (« mappings », contrôles, « workflows », constats, imports, exports etc.).

ET APRÈS ?

NumaHOP, dont le code source (<https://github.com/progilone/numahop>) est sur Github, résulte d'une démarche résolument engagée en faveur de l'utilisation et du partage d'applications open source. La prise en main de cet outil par une communauté d'utilisateurs sera décisive pour son développement et son évolution. Une diffusion large de son utilisation au sein de la communauté universitaire et de toutes les institutions concernées engagerait une démarche partagée d'améliorations et d'harmonisation des pratiques, en favorisant la mutualisation des savoir-faire et des connaissances.

↖ Périimètre fonctionnel de NumaHOP

OLESEA DUBOIS

Responsable service Numérisation et Archivage numérique à la Bibliothèque de Sciences Po
olesea.dubois@sciencespo.fr

FANNY MION-MOUTON

Responsable adjointe du pôle Flux et données à la Bulac
fanny.mion-mouton@bulac.fr

PAULINE RIVIÈRE

Chef de projet numérisation à la Bibliothèque Sainte-Geneviève
pauline.riviere@sorbonne-nouvelle.fr