

NumaHop, plate-forme de gestion des contenus numérisés

Olesea Dubois

Bibliothèque de Sciences Po
30 Rue Saint-Guillaume
75007 Paris

Fanny Mion-Mouton

BULAC
65 Rue des Grands Moulins
75013 PARIS

Pauline Rivière

Bibliothèque Sainte-Geneviève
10 place du Panthéon
75005 PARIS

Résumé

NumaHop est un logiciel libre (licence GNU Affero General Public License v3.0) mis en œuvre par la ComUE Université Sorbonne Paris Cité avec le soutien financier de la ville de Paris. Il a été développé sous la direction des établissements pilotes : la bibliothèque Sainte-Geneviève, la bibliothèque de Sciences Po Paris et la BULAC (bibliothèque universitaire des langues et civilisations). Il est librement réutilisable par tout établissement et son code source est en ligne sur GitHub. NumaHop est un outil intégré qui permet de gérer une chaîne de numérisation de documents 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 entre les différentes étapes de la numérisation impliquant les acteurs concernés : ABES (Agence bibliographique de l'enseignement supérieur), prestataires de numérisation, bibliothèques, DSI, diffuseurs, CINES (Centre Informatique National de l'Enseignement Supérieur).

Mots-clefs

Numérisation, Chaîne de numérisation, Bibliothèques numériques, Métadonnées, Archivage pérenne.

1 Introduction

NumaHop est un projet porté par trois établissements pilotes (bibliothèque Sainte-Genève, bibliothèque de Sciences Po-Paris, BULAC), dans le cadre de la Comue Sorbonne Paris Cité, et avec le soutien financier de la Ville 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.

2 Présentation générale

2.1 Objectifs du projet et Méthodologie de travail

L'objectif attendu était triple.

En premier lieu, le logiciel développé devait être open source¹ pour être diffusé, utilisé et amélioré le plus largement possible par l'ensemble des établissements intéressés, dans la sphère universitaire et culturelle (bibliothèques, archives, musées), mais aussi au-delà (associations, entreprises).

Par ailleurs, la plate-forme de gestion des contenus numérisés devait 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 pour favoriser l'appropriation par l'ensemble des agents, y compris sur des tâches habituellement identifiées comme "techniques".

Plusieurs ateliers menés en lien avec la société Progilone tout au long de l'année 2016-2017 ont permis d'aboutir à une version POC (*Proof of Concept*) en février 2017, puis à la livraison d'une version finale en juin 2018, avant un passage en production dans les trois établissements pilotes à la fin de l'année 2018 et au début de l'année 2019.

1. Code source disponible en ligne sur [GitHub](https://github.com/progilone/). Licence GNU Affero General Public License v3.0 : <https://github.com/progilone/>

Désormais utilisé au quotidien par les équipes, NumaHop a permis de constater rapidement une simplification des chaînes de numérisation de documents. Cette plateforme, accessible en full web permet aujourd’hui 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, DSI, diffuseurs, CINES).

2.2 Modules fonctionnels de NumaHop

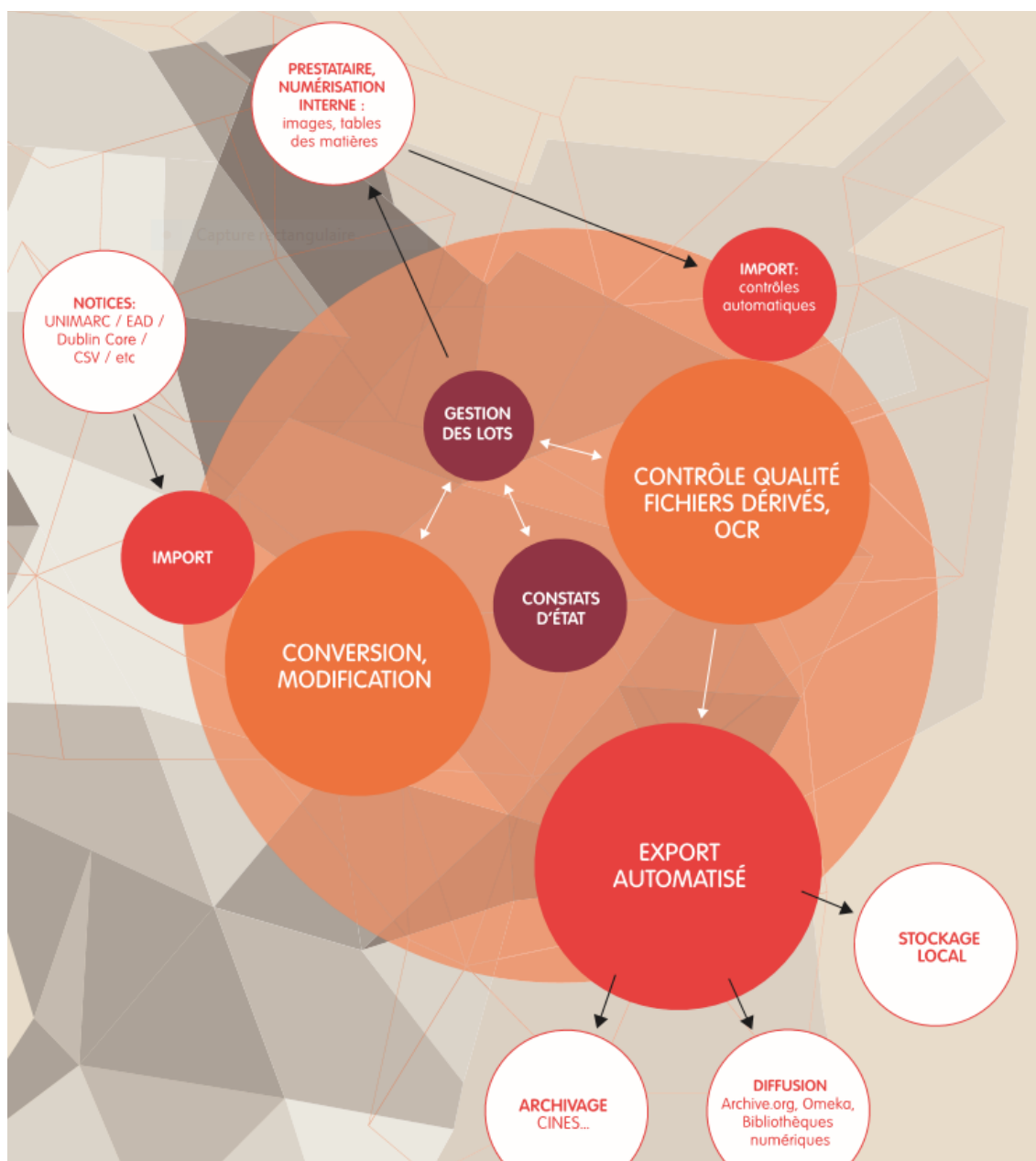


Illustration 1: Périmètre fonctionnel de NumaHop

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 EAD dans des formats interopérables : *Dublin Core*, *Dublin Core* qualifié. Les mappings sont personnalisables et adaptables selon les spécificités de chaque établissement. Ce module permet également d'importer les fichiers images correspondant aux vues numériques, qu'elles soient livrées par un prestataire de numérisation ou directement produites par l'établissement lui-même. 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, de nommage (séquençage, casse...).

Fiche
Livrer

Date de réception [30/01/2019](#)

Dossier de livraison [livraisonBULAC](#)

Notes de livraison [Non renseigné](#)

Notes de contrôle livraison [Non renseigné](#)

Echantillonnage [Pas d'échantillonnage](#)

▼ Documents numériques rattachés 8

Nom	Date livraison	Nb. pages	Poids	Statut	Contrôles	Action
BULAC_MS_ARA_25	30/01/2019	53	138,25 Mo	En cours de contrôle	●	
BULAC_MS_ARA_39	30/01/2019	193	316,09 Mo	En cours de contrôle	●	
BULAC_MS_ARA_4	30/01/2019	0	- - -	Rejeté	●	
BULAC_MS_PERS_33	30/01/2019	395	862,39 Mo	En cours de contrôle	●	
BULAC_MS_PERS_55	30/01/2019	181	243,18 Mo	A contrôler	●	
BULAC_MS_PERS_75b	30/01/2019	115	568,16 Mo	En cours de contrôle	●	
BULAC_MS_PERS_80	30/01/2019	109	218,34 Mo	En cours de contrôle	●	
BULAC_MS_PERS_93	30/01/2019	309	441,45 Mo	En cours de contrôle	●	

Charger le rapport

Illustration 2: Écran de suivi des livraisons dans NumaHop


Fiche ✎ 📄 🗑️

Taux d'erreurs mineures [0,00%](#)

Taux d'échantillonnage [0,00%](#)

Méthode d'échantillonnage [Pas d'échantillonnage](#)

Séparateur de séquence [-](#)

Bibliothèque [Bibliothèque Sainte-Geneviève](#) 

Activé	Bloquant	Contrôle
<input checked="" type="checkbox"/>		Présence de fichiers masters
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle du format
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle de la séquence
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle du nombre total de fichiers
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle du type de compression
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle du taux de compression
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle de la résolution
<input type="checkbox"/> Désactivé	<input type="checkbox"/> Non	Contrôle du profil de couleurs
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle nom de fichier préfixé par le préfixe de bibliothèque
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle de la casse du nom de fichier
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle de la table des matières
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Contrôle présence fichier pdf multicouches
<input checked="" type="checkbox"/>	<input type="checkbox"/> Non	Validation FACILE par le Cines

Illustration 3: Écran de paramétrage des contrôles automatisés d'images réalisés lors de la livraison dans NumaHop

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. Divers outils de gestion de projets (statistiques, tableaux de bord...) sont associés à ce module, pour faciliter le recueil d'indicateurs d'activités pertinents.

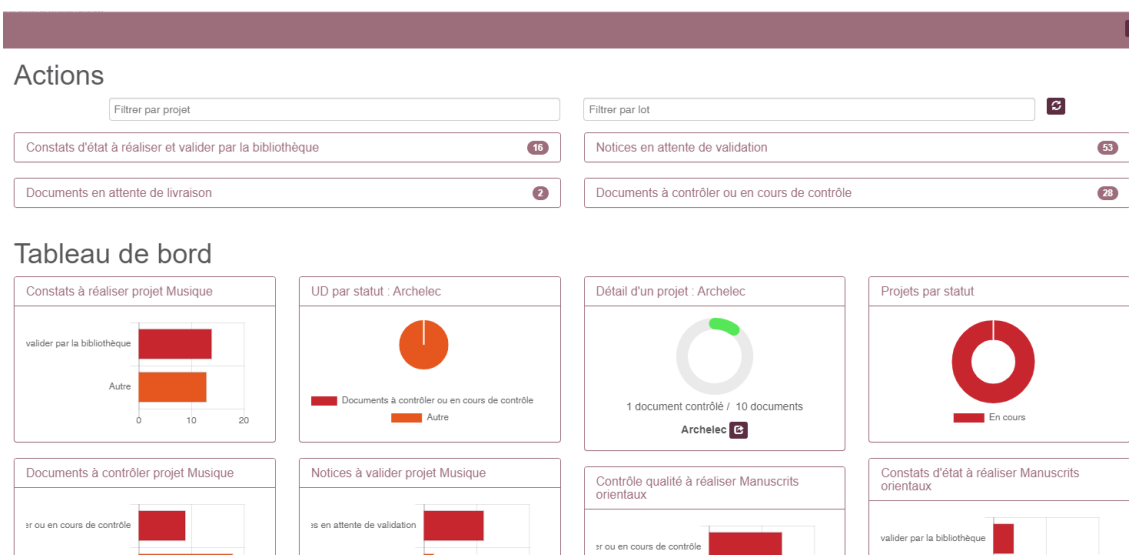


Illustration 4: Tableau de bord de NumaHop

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, OCR...), 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 en particulier à ses performances. La création, en temps réel, par le logiciel, de fichiers dérivés d'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.

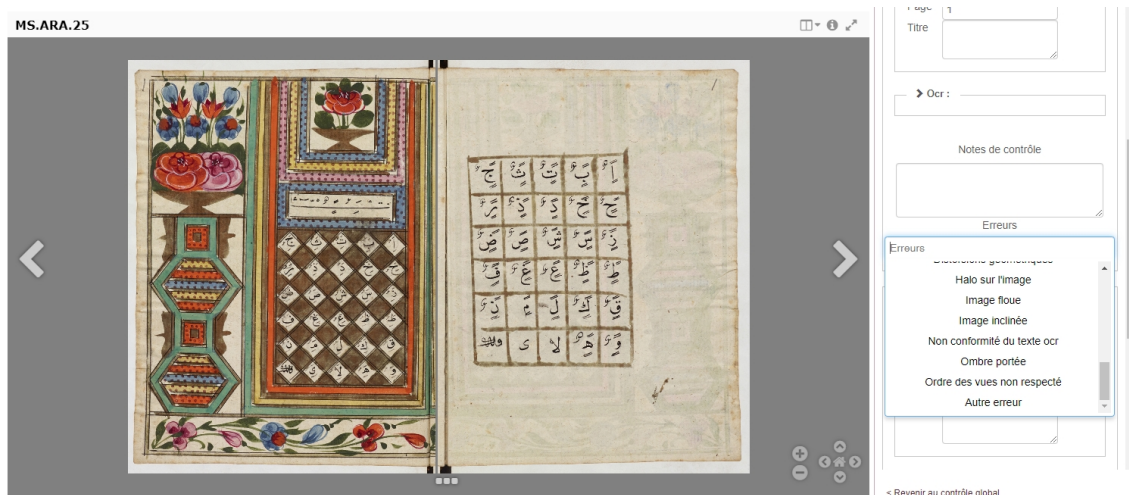


Illustration 5: Interface de contrôle qualité

Enfin, le dernier module concerne l'ensemble des 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 (serveur interne de l'établissement) 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 les fichiers sip.xml et METS 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, auparavant très chronophages et fastidieuses, les établissements peuvent disséminer largement et de manière systématique l'ensemble de leurs contenus numérisés.

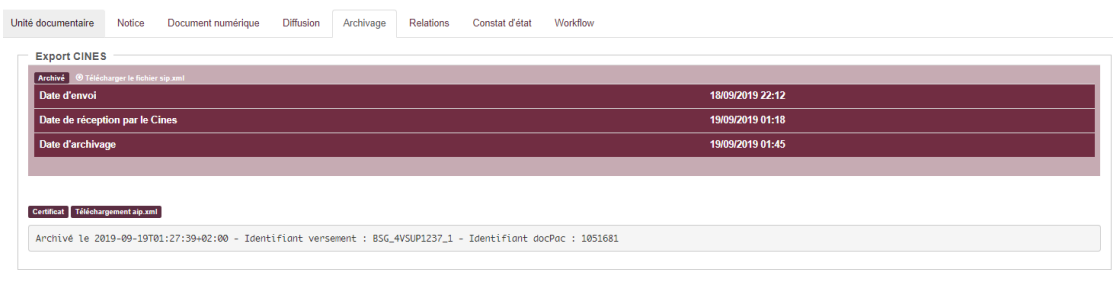


Illustration 6: Écran de contrôle de l'archivage automatisé

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

2.3 Avantages techniques, financiers et organisationnels engendrés par l'utilisation de NumaHop

L'utilisation de NumaHop pour la gestion des projets de numérisation a mis en valeur de nombreux avantages techniques, financiers, mais aussi organisationnels. L'interface ergonomique permet une acculturation rapide des personnels, et l'utilisation de méthodes de travail standardisées garantit une normalisation des processus. La création automatisée par l'outil d'un certain nombre de fichiers entraîne une diminution des coûts associés à la numérisation. L'utilisation d'un outil unique simplifie les interactions avec le prestataire choisi (contrôle qualité fiabilisé, envoi de bordereaux normalisés...) et le suivi de projet au quotidien. Enfin, grâce à la diffusion de ce nouvel outil au sein de la communauté universitaire et de toutes les institutions concernées, les établissements pilotes espèrent développer une communauté d'utilisateurs, dans une démarche partagée d'améliorations et d'harmonisation des pratiques, en favorisant la mutualisation des savoir-faire et des connaissances.