olkoa

ETUDE SUR LA PRESERVATION DES COLLECTIONS DES BIBLIOTHEQUES NUMERIQUES DE L'ESR

VERS LA CONSTITUTION D'UNE COMMUNAUTE DE PRATIQUES

CollEx-Persée

Table des matières

1. PREAMBULE 3
2. SYNTHESE DES PRECONISATIONS 5
3. CONTEXTE ET METHODOLOGIE 6
3.1 CONTEXTE 6
3.2 METHODOLOGIE 7
4. METTRE EN PLACE UNE GOUVERNANCE ADAPTEE AUX ENJEUX : ELEMENTS DE REFLEXION 9
4.1 ENCOURAGER LA MONTEE EN COMPETENCE DES ETABLISSEMENTS SUR LE SUJET DE LA PRESERVATION 9
4.2 CONSTRUIRE UN CADRE FEDERATEUR 10
4.2.1 PRECONISATIONS 11
5. FAIRE CONVERGER LES PRATIQUES DOCUMENTAIRES 12
5.1 ETAT DES LIEUX 12
5.1.1 CONTEXTES DE PRODUCTION DES METADONNEES 12
5.1.2 CARACTERISTIQUES DES FICHIERS METS PRODUITS PAR LES ETABLISSEMENTS 12
5.1.3 MODALITES DE PRESERVATION DES COLLECTIONS 13
5.1.4 CONSTATS 14
5.2 MODALITES PROPOSEES 14
5.2.1 ENRICHIR LA DESCRIPTION DES COLLECTIONS NUMERISEES 14
5.2.2 NORMALISER LA DESCRIPTION DES COLLECTIONS NUMERISEES 15
5.2.3 PRECONISATIONS 16
6. ANNEXES 18
6.1 GROUPE DE TRAVAIL PRESERVATION 18
6.2 CONVENTION DE FORMATION AVEC L'ENSSIB 19
6.3 LETTRE DE MISSION 23
6.4 CARACTERISTIQUES DES COLLECTIONS NUMERISEES : ETAT DES LIEUX 25

1. Préambule

Le GIS CollEx-Persée, en tant qu'infrastructure de recherche en information scientifique et technique, encourage et soutient la constitution de corpus et collections numériques qui viennent compléter, enrichir ou documenter les collections des établissements de l'Enseignement supérieur et de la recherche (ESR). Pour autant, la question de la préservation reste un « angle mort » des politiques des établissements qui, pour nombre d'entre eux, n'ont pas mis en œuvre de solution d'archivage pérenne. C'est pourtant un impératif pour garantir que ce patrimoine numérique ne soit pas perdu – et que les efforts financiers consentis pour le constituer n'aient pas été réalisés en vain.

De quoi parle-t-on ? La préservation numérique se définit comme une série d'activités pour garantir un accès continu aux objets numériques aussi longtemps que nécessaire. Elle englobe toutes les actions requises de manière à maintenir l'accès aux objets numériques en dépit de la défaillance des équipements ou des changements ayant lieu au niveau technologique et organisationnel[1]. Elle présuppose d'avoir suffisamment d'informations contextuelles sur les données (métadonnées) pour pouvoir les préserver et pour que la communauté d'utilisateurs puisse évaluer leur qualité, c'est-à-dire leur authenticité et intégrité.

Pour de nombreux établissements, la préservation numérique est entièrement nouvelle et les ressources adéquates doivent être rassemblées pour la première fois. C'est pourquoi la capacité à échanger avec d'autres établissements et d'autres organisations est devenue une compétence clé pour les gestionnaires de la préservation numérique. La sensibilisation, l'acquisition et le partage de connaissance doivent être une préoccupation permanente plutôt que ponctuelle.

Historique des travaux. Dès 2019, le groupe de travail « Préservation numérique » de CollEx-Persée[2] a cherché à comprendre les causes de cette prise en compte insuffisante et à proposer des solutions pour y remédier. Après s'être intéressé à la question des contenus à préserver, puis aux enjeux de formation, il s'est penché sur les questions financières. Certes, les solutions et les prestataires existent, mais les coûts sont encore considérables. Pour diminuer les tarifs, et pour encourager les établissements à « sauter le pas » de la préservation, le groupe de travail a considéré que la meilleure méthode était de réaliser des économies d'échelle. En effet, le prix au « téraoctet » facturé par les prestataires d'archivage diminue en fonction du volume total archivé ; en réunissant les collections de plusieurs établissements pour n'en faire qu'une, on arriverait à des coûts unitaires plus soutenables.

L'appel à manifestation d'Intérêt de mai 2021 a réuni différentes institutions membres du réseau pour constituer un groupement d'achat commun. Cet AMI invitait chaque candidat à présenter ses collections numériques (intérêt pour la recherche et caractéristiques techniques : volume, formats de données et de métadonnées) et indiquer ses besoins en matière de préservation. Sept établissements ont postulé et ont été retenus pour participer à l'élaboration d'une solution commune. En parallèle, un groupe d'experts comprenant des représentants de la Bibliothèque nationale et universitaire, des bibliothèques de Sciences-Po et du Muséum national d'histoire naturelle, et du Service

[1] Voir notamment les travaux de la DPC - Digital Preservation Coalition : https://www.dpconline.org/about

[2] Page du groupe : https://www.collexpersee.eu/groupes/preservation-numerique/

interministériel des Archives de France, puis de l'UAR Persée, a élaboré un programme de travail à leur intention.

Une première décision stratégique a été de se concentrer sur la question des collections numérisées (qu'il s'agisse de textes, d'archives ou d'images) diffusées par les bibliothèques numériques des institutions. Il s'agit, pour les établissements, de l'ensemble de documents le plus volumineux (en téraoctets), mais aussi celui pour qui les procédures de production, de gestion et de diffusion, ainsi que de préservation, sont les plus mûres. Les établissements retenus pour l'AMI ont également indiqué disposer de documents nativement numériques (images, bureautique, audiovisuel, etc.) mais il s'agit d'un enjeu moins immédiatement critique, du moins d'un point de vue volumétrique et donc financier.

CollEx-Persée a alors engagé une démarche de pédagogie et de sensibilisation avec ces sept établissements en vue d'identifier leurs besoins et problématiques et entamer un dialogue spécifique avec les opérateurs compétents en matière de préservation numérique. L'enjeu est de constituer, à l'échelle des établissements, une communauté de pratiques en vue de parvenir à un dispositif (organisationnel et technique) commun et partagé permettant de préserver le patrimoine documentaire que représentent les collections des bibliothèques numériques des établissements de l'ESR.

L'action du programme s'est développée entre mai 2022 et juin 2023 selon deux volets successifs :

- Tout d'abord, les participants à l'AMI ont bénéficié d'une formation afin d'harmoniser leurs compétences sur le sujet, et adopter un vocabulaire commun. Pour cette formation, le GIS CollEx-Persée a sollicité le soutien de l'Enssib, qui a accepté de s'y associer,
- Ensuite, ils ont participé à une réflexion commune, animée par un consultant en dématérialisation et conservation numérique, pour faire converger leurs pratiques documentaires de sorte que leurs collections numérisées soient suffisamment similaires pour être considérées, par les acteurs de la préservation, comme une collection unique.

En résumé – **Les objectifs étaient les suivants** :

- Faire émerger la préservation comme sujet de discussion collective dans l'ensemble des établissements de l'ESR concernés grâce à leur montée en compétences sur le sujet,
- Encourager la modernisation et la convergence des pratiques et la définition d'une chaîne commune de production, description, accès et préservation des collections numérisées. C'est cette convergence qui autorise à envisager des économies d'échelle grâce à la mise en œuvre d'une filière de préservation commune à l'ensemble des établissements de l'ESR concernés, prérequis à la mutualisation de l'archivage pérenne,
- Approfondir la question de la coordination de la préservation des collections numérisées et plus généralement, numériques pour l'ensemble des établissements de l'ESR concernés.

L'objet du présent rapport est de décrire les conditions pour faire émerger et pérenniser cette communauté de pratiques afin de soutenir la préservation des collections des bibliothèques numériques de l'enseignement supérieur et de la recherche. Il formule des orientations stratégiques et reflète un état de la réflexion qui reste à poursuivre dans le cadre du GIS CollEx Persée 2.

2. Synthèse des préconisations

Faire émerger une communauté de pratiques

- Reconduire et pérenniser le programme de formation mis en place, en 2023, par CollEx- Persée avec l'Enssib.

Participer à la définition commune d'une stratégie de préservation

- Intégrer CollEx-Persée comme un partenaire d'un futur dispositif national de préservation des collections numérisées des établissements de l'ESR.
- Préciser le rôle et la fonction de CollEx-Persée par rapport aux opérateurs nationaux sous tutelle du MESR et impliqués dans ce domaine, particulièrement le CINES, l'Abes et l'UAR Persée.

Faire converger les pratiques documentaires

- Réorienter les travaux du groupe de travail Préservation vers la production d'un set de métadonnées commun pour la préservation des collections numérisées.
- Définir un profil METS des établissements de l'ESR, à partir de ce travail, et en lien avec tous les opérateurs concernés – CINES, Abes et UAR Persée.
- Travailler avec la communauté des utilisateurs de NumaHOP pour enrichir le fichier METS produit par cette application, et faire financer les développements dans le cadre du dispositif CollEx-Persée, après une réflexion en coordination avec l'Abes et l'UAR Persée afin que le fichier produit soit conforme au profil METS souhaité.
- Dans un souci d'interopérabilité, faire en sorte que NumaHOP puisse aussi produire un fichier de métadonnées au format SEDA en complément ou en remplacement du format METS, comprenant les mêmes métadonnées (par exemple en transformant, via une feuille de style, le fichier METS au format SEDA).
- Dans un souci de traçabilité, promouvoir la mise en œuvre de PGD permettant de documenter les choix (et leur évolutions) opérés par les établissements et d'alimenter les métadonnées du METS.
- En parallèle, étudier et évaluer les autres outils utilisés par les bibliothèques de l'ESR (chaîne Persée, collaborations avec Huma-Num) pour assurer la pérennité des données qu'elles produisent notamment lorsque la numérisation s'accompagnement d'enrichissements.

3. Contexte et méthodologie

3.1 Contexte

Le groupe de travail « Préservation numérique » de CollEx-Persée rassemble les principaux acteurs de la préservation numérique et les représentants de bibliothèques déjà impliquées sur le sujet : le Ministère de l'enseignement supérieur et de la recherche, le Centre informatique national de l'enseignement supérieur (CINES), la Bibliothèque nationale de France (BnF), l'infrastructure de recherche Huma-Num, le Service interministériel des archives de France (SIAF), l'Agence bibliographique de l'enseignement supérieur (Abes), l'UAR Persée, Sciences Po Paris, La Contemporaine, le Muséum national d'histoire naturelle et la Bibliothèque nationale et universitaire, ces deux dernières institutions étant co-pilotes du groupe. En mai 2021, le groupe de travail a lancé un appel à manifestation d'intérêt (AMI)[3] auprès des membres du réseau CollEx-Persée pour apporter une réponse commune aux enjeux de préservation des collections numérisées. Sept établissements ont été retenus : la Bibliothèque interuniversitaire de la Sorbonne, la Bibliothèque nationale et universitaire, la Bibliothèque Sainte-Geneviève, la Bibliothèque de l'Institut français de Pondichéry, l'Institut national d'histoire de l'art, le SCD de Lyon 1 et la Bibliothèque du Muséum national d'histoire naturelle.

Les établissements membres de l'AMI ont suivi un cycle de formation coorganisé avec l'École nationale supérieure des sciences de l'information et des bibliothèques (Enssib), en mai et juin 2022[4].

Un appel d'offres a été lancé en septembre 2022[5], par CollEx-Persée, pour accompagner les établissements membres de l'AMI dans leur réflexion sur la mutualisation de la préservation de leurs collections numérisées et la constitution d'un groupement de commandes.

L'étude confiée à la société OLKOA à l'automne 2022 a pour objectifs de :

- S'assurer que les collections numérisées disposent des métadonnées utiles et nécessaires à leur préservation,
- Formuler des recommandations pour converger vers des pratiques de description des collections des bibliothèques numériques homogènes afin d'envisager des économies d'échelle lors des négociations collectives avec les acteurs nationaux de la préservation.

L'étude s'est déroulée entre novembre 2022 et septembre 2023.

[3] https://www.collexpersee.eu/groupes/preservation-numerique/
[4] Voir en annexe la convention de formation.
[5] Voir en annexe la lettre de mission pour l'accompagnement pour la constitution d'un groupement de commandes en vue d'une prestation de préservation numérique.

3.2 METHODOLOGIE

L'étude s'est déroulée en deux phases :

- Une première phase d'état des lieux afin d'identifier les convergences et les différences dans les pratiques documentaires de description des collections numérisées et formuler des recommandations,
- Une seconde phase d'évaluation de la faisabilité de l'archivage des collections des bibliothèques numériques des établissements de l'ESR auprès de la BnF ou du CINES.

La première phase de l'étude a permis de dresser un état des lieux des caractéristiques des collections numérisées et des pratiques documentaires liées à leur description[6]. L'objectif était d'une part d'évaluer la qualité des métadonnées disponibles (standards de description utilisés, exhaustivité, cohérence, nature et caractéristiques des métadonnées) et, d'autre part, de faire des recommandations pour tendre vers un alignement des métadonnées, en vue de la production de métadonnées homogènes.

Cet état des lieux s'est appuyé sur :

- Des entretiens avec les établissements ayant répondu à l'AMI, le 8 novembre 2022,
- Un échange avec les équipes de NumaHOP, le 24 novembre 2022,
- La mise à disposition d'un panel de données jugées représentatives par chacun des établissements.

Cette première phase de l'étude a permis de formuler un ensemble de préconisations reprises dans le chapitre 4 du présent rapport.

La seconde phase de l'étude avait pour objectif d'analyser les opportunités, conditions et prérequis à la mutualisation de la préservation des collections des bibliothèques numériques des établissements, en s'appuyant sur l'état de des lieux réalisé précédemment sur les collections des établissements ayant répondu à l'AMI. Il était prévu d'organiser des entretiens avec la BnF et le CINES ; ces entretiens devaient permettre, d'une part, de s'assurer que le postulat de l'homogénéisation des collections était bien un levier de négociation pour faire des économies d'échelle et, d'autre part, d'identifier les conditions à respecter et à mettre en œuvre pour que les collections des bibliothèques numériques puissent être traitées comme une collection unique. En cours d'étude, l'objet de cette seconde phase a été réorienté au profit de la fourniture d'un ensemble de recommandations, objet du présent rapport, en vue de la constitution d'une communauté de pratiques à l'échelle des bibliothèques de l'ESR. Pour autant, il a été décidé de maintenir une coopération technique avec la BnF afin de s'inscrire dans la convergence des pratiques de description des collections des bibliothèques numériques.

Au cours de cette seconde phase, les entretiens suivants ont été menés :

- Atelier avec les établissements membres de l'AMI pour consolider leurs besoins de préservation et le niveau de service attendu, le 24 janvier 2023,
- Echange entre Olkoa et la BULAC, à savoir Nathalie Pilet, chargée de la bibliothèque numérique et des thèses, sur la méthodologie de reprise des données

[6] Voir en annexe.

avec NumaHOP et leur transfert dans la plateforme PAC du CINES, le 30 janvier 2023,

- Echange entre Olkoa et le CINES, à savoir Aulsen Dzangui, e-archiviste au Département Archivage et Service aux Données, le 20 mars 2023, pour présenter la réflexion portée par CollEx-Persée,
- Echange avec des représentants des établissements et la BnF, à savoir Sébastien Pétratos, directeur du Département images et prestations numériques et Thomas Ledoux responsable de la mission de coordination des entrées, le 29 mars 2023, pour présenter la réflexion portée par CollEx-Persée,
- Echange avec des représentants des établissements et l'Abes, à savoir Nicolas Morin, directeur, Marianne Giloux, responsable du Département Métadonnées et services aux Réseaux et Stéphane Gully, DSI, le 29 mars 2023,
- Atelier technique entre Olkoa et les experts METS de la BnF, Thomas Ledoux et Bertrand Caron, expert formats le 26 avril 2023,
- Echange tripartite entre le CINES, à savoir Philippe Prat du département archivage et données et Aulsen Dzangui e-archiviste, l'Abes, à savoir Nicolas Morin, et CollEx, le 22 septembre 2023.

Cette seconde phase de l'étude a permis de faire émerger les conditions d'une collaboration solide entre les établissements et acteurs de la communauté de l'ESR, pour passer à l'échelle.

Le chapitre suivant du présent rapport revient sur le cadre de gouvernance.

4. Mettre en place une gouvernance adaptée aux enjeux : éléments de réflexion

Lorsqu'elle existe, la stratégie de préservation des collections numérisées est pensée et portée à l'échelle de chaque établissement. Le souhait serait d'harmoniser cette approche avec une échelle plus collective propre au réseau en travaillant en amont de cette réflexion. Les établissements qui se sont engagés dans une politique de préservation de leurs collections numérisées ont fait le choix, dicté par leur contexte, de faire appel soit au CINES, soit à la BnF comme opérateur de services de préservation.

L'une des ambitions de l'AMI était, pour favoriser le passage à l'échelle, de constituer un groupement de commande pour la fourniture d'un service de préservation des collections des bibliothèques numériques des établissements. L'étude a permis de mettre en avant les conditions et prérequis à la mutualisation de ce service de préservation avec d'une part la standardisation des métadonnées et d'autre part l'importance du pilotage du projet en interface avec le prestataire de préservation. Pour réunir ces deux conditions, il est un préalable, celui de la coopération entre les établissements.

4.1 Encourager la montee en competence des etablissements sur le sujet de la preservation

L'AMI Préservation a favorisé la coopération entre les établissements participants sur le sujet de la préservation numérique avec :

- La mobilisation et la circulation des expertises notamment concernant NumaHOP,
- La montée en compétences des établissements sur le sujet de la préservation numérique avec l'organisation d'un cycle de formations, dans le cadre d'une convention de formation professionnelle entre CollEx-Persée et l'Enssib.

Un parcours de formation a été proposé aux établissements membres de l'AMI avec : un premier niveau de sensibilisation à distance avec 2 webinaires de 3h ayant comme objectif de fournir un éclairage commun sur les pratiques et les enjeux de la préservation numérique et un second niveau de perfectionnement d'une durée de 12h en présentiel et 3h à distance sur la chaîne de traitement d'une opération d'archivage pérenne. Ce parcours de formation est désormais inscrit au catalogue de l'Enssib afin de servir au-delà des membres du présent AMI, ce qui était un des objectifs visés par CollEx-Persée dans cette démarche.

La formation des établissements aux enjeux de la préservation mais aussi aux aspects plus techniques liés à la description des documents numériques et aux formats de métadonnées, et notamment au format METS, est un incontournable.

La coopération entre les établissements doit être poursuivie et le cercle de la coopération élargi pour favoriser et encourager la montée en compétences des établissements sur le sujet de la préservation numérique. La sensibilisation, l'acquisition

et le partage de connaissance doivent être une préoccupation permanente plutôt que ponctuelle.

4.2 CONSTRUIRE UN CADRE FEDERATEUR

Le second axe de l'étude a permis de réfléchir à la définition d'un service commun de préservation, dans le cadre d'une stratégie de préservation des collections numérisées des bibliothèques de l'ESR.

Il a confirmé que l'éventail des compétences requises pour surmonter efficacement les défis de la préservation numérique nécessite de rechercher une plus grande coopération entre les acteurs, partenaires institutionnels et opérateurs de l'ESR et d'inscrire cette coopération dans le cadre d'une gouvernance partenariale pérenne.

Le CINES est l'opérateur cible pour les établissements de l'ESR ; deux établissements membres de l'AMI ont d'ailleurs déjà contractualisé avec le CINES. L'harmonisation des pratiques documentaires de description et l'homogénéisation des métadonnées doivent permettre des économies d'échelle en considérant les collections numérisées des bibliothèques comme une unique collection. Cette convergence des pratiques de description doit s'appuyer sur la définition d'un schéma national pour coordonner les différents acteurs et assurer l'interface entre les établissements et le CINES. Il a été proposé, lors de l'entretien du 22 septembre avec le CINES et l'Abes, de travailler ensemble à cette harmonisation, en lien également avec l'UAR Persée.

La coopération suppose une politique partenariale entre les acteurs et la co-construction d'un cadre fédérateur au dialogue entre les établissements, l'Abes et le CINES. Le CINES et l'Abes ont prévu de réfléchir aux modalités de cette coordination nationale. Ils souhaitent, cependant, prendre un temps supplémentaire pour évaluer les moyens humains et financiers nécessaires dans la longue durée pour instruire ce dossier et se tourner vers le MESR avant de revenir auprès de la communauté CollEx-Persée. L'Abes a indiqué ne pas l'inscrire dans son actuel projet d'établissement mais rester en veille sur ce sujet.

La cartographie des acteurs et les scenarii de coopération restent donc à consolider mais les attentes identifiées et partagées par les établissements membres de l'AMI peuvent être formulées de la manière suivante :

- Être associés au dialogue avec le MESR, l'Abes et le CINES pour la définition des orientations stratégiques,

Dans ce contexte, CollEx-Persée pourrait participer à une réflexion globale sur la mise en place d'un dispositif mutualisé de préservation des collections, en offrant aux établissements intéressés un espace de dialogue ; l'Abes étant dans tous les cas l'interface des établissements vis-à-vis du CINES. Quant aux liens avec la BnF, ils doivent être entretenus pour favoriser la convergence des pratiques.

En outre, pour ce qui concerne les développements à réaliser sur NumaHOP, et qui sont détaillés dans le chapitre suivants, un dialogue doit être instauré avec la communauté des utilisateurs de NumaHOP, qui est en train de s'organiser dans le cadre du projet de l'Ouvroir 2 financé par Collex-Persée.

Selon la feuille de route soumise au Conseil des membres en décembre 2022, le futur GIS devra se structurer en grands programmes : cartographie/labellisation, numérisation enrichie, acquisitions électroniques, archives scientifiques. Les enjeux de la préservation numérique ont toute leur place dans le futur GIS, comme cela est souligné

par le récent rapport de l'IGESR sur CollEx-Persée[7]. Sa place est cependant à déterminer : sans doute comme une composante d'un ou plusieurs programmes, comme numérisation enrichie (pour les données des bibliothèques numériques) et archives scientifiques (pour les archives nativement numériques) ? On ne peut exclure, non plus, la création d'un groupe de travail directement rattaché au porteur du GIS ou un programme blanc développé en parallèle. Cela demandera une instruction fine avec l'ensemble des établissements impliqués. Quelle que soit la solution retenue, il importe que l'on puisse capitaliser sur le travail mené de façon exploratoire par les membres de l'AMI Préservation afin de constituer une communauté la plus large possible.

Notons enfin que ces réflexions concernent la préservation des collections numériques, mais pas leur accès à destination du public. Or, cette préservation n'a pas de sens sans le maintien d'un accès public aux collections, qui doit se faire par les biais de bibliothèques numériques. À ce titre, des liens doivent être entretenus entre le groupe préservation numérique de CollEx-Persée et l'opérateur Persée, mais aussi avec d'autres acteurs de la communauté des bibliothèques numériques (fournisseurs de plateformes, communautés de logiciels *open source*…). Si les établissements participants au sein du réseau CollEx-Persée parviennent à fournir des documents dotés de formats permettant leur pérennisation, la discussion entre tous ces acteurs s'en trouvera facilitée.

4.2.1 Préconisations

Reconduire et pérenniser le programme de formation mis en place, en 2023, par CollEx-Persée avec l'Enssib.

Intégrer CollEx-Persée comme un partenaire d'un futur dispositif national de préservation des collections numérisées des établissements de l'ESR.

Préciser le rôle et la fonction de CollEx-Persée par rapport aux opérateurs nationaux sous tutelle du MESR et impliqués dans ce domaine, particulièrement le CINES, l'Abes et l'UAR Persée.

[7]https://www.collexpersee.eu/rapport-de-ligesr/, p. 13-14

5. Faire converger les pratiques documentaires

La première phase de l'étude, sur l'état des lieux des collections numérisées, a porté à la fois sur les caractéristiques des collections numérisées et sur un diagnostic de la qualité des métadonnées. Cet état des lieux a permis d'identifier les convergences et divergences des pratiques documentaires des établissements membres de l'AMI. L'attention s'est portée sur les pratiques liées à la description des collections numérisées et notamment la nature et la qualité des métadonnées.

5.1 ETAT DES LIEUX

5.1.1 Contextes de production des métadonnées

Les métadonnées sont indispensables pour retrouver un contenu et préserver les collections. Elles décrivent le contenu lui-même (métadonnées descriptives), les fichiers informatiques issus des opérations de numérisation (métadonnées techniques), l'historique de production du contenu et des fichiers (métadonnées de provenance), l'organisation et la relation des fichiers entre eux pour représenter le contenu (métadonnées de structure). Différents standards définissent leur syntaxe et leur forme.

Les caractéristiques des métadonnées utilisées pour décrire les collections numérisées sont les suivantes :

- Les métadonnées descriptives sont extraites des catalogues CALAMES ou SUDOC ; elles sont importées ou renseignées directement dans les interfaces des bibliothèques numériques,
- Les métadonnées de gestion, lorsqu'elles existent, sont produites au cours de la chaîne de numérisation via NumaHOP, par les prestataires de numérisation ou via le logiciel OMEKA de gestion de bibliothèques numériques ; elles sont structurées selon le schéma XML METS ; quatre établissements sur six produisent des fichiers METS.

5.1.2 Caractéristiques des fichiers METS produits par les établissements

Un fichier METS comprend plusieurs sections et principalement :

- Les métadonnées d'identification et de description des objets numérisés ou métadonnées descriptives,
- Les métadonnées de production qui portent sur l'aspect technique de la numérisation et sur son historique ou métadonnées administratives,
- Les métadonnées relatives aux fichiers images ou métadonnées techniques,
- La carte des structures avec les métadonnées d'empaquetage qui permettent de rattacher les métadonnées à un document ou à une page du document, par exemple.

Les fichiers METS produits par les établissements ont les caractéristiques suivantes :

- Les métadonnées descriptives (section dmdSec) sont exprimées en Dublin Core et encapsulées dans le fichier METS ; Le lien vers la notice MARC dans le SUDOC est parfois matérialisé,
- Les métadonnées relatives aux opérations de numérisation (section amdSec) sont pauvres. Lorsqu'elles existent, elles sont exprimées selon le schéma NISO (ANSI/NISO Z39.87-2006 (R2017) Data Dictionary - Technical Metadata for Digital Still Images),
- Les métadonnées techniques (section fileSec) comprennent *a minima* l'indication du Type MIME du fichier ; l'empreinte du fichier n'est pas systématiquement indiquée,
- La carte des structures (structMAP) définit la structure hiérarchique de l'objet numérique décrit ; cette section est systématiquement implémentée.

5.1.3 Modalités de préservation des collections

Trois établissements préservent leurs collections auprès d'un tiers, la BnF pour un établissement et le CINES pour deux établissements. Les autres établissements n'ont pas encore de politique de préservation : leurs collections numérisées sont stockées sur des supports amovibles ou sur serveurs exploités par les DSI.

Pour les établissements qui confient la préservation de leurs collections numériques à un tiers, les formats et la structuration des métadonnées descriptives et de gestion sont ceux requis par le prestataire de préservation :

- Au CINES, le fichier METS n'est pas exploité. En effet, le schéma de métadonnées implémenté sur leur plateforme d'archivage électronique (PAC) est le SEDA (Standard d'échange de données pour l'archivage). Le ficher METS est traité comme un fichier de données à préserver, au même titre que les fichiers images même s'il n'est pas inclus dans le paquet d'archives (SIP) ; il n'est pas contrôlé ni validé et sa présence n'est pas obligatoire ; sa structuration et son contenu (valeurs) ne sont pas imposées par le CINES ; en revanche, lorsqu'il existe, il est décrit dans le manifeste sip.xml qui accompagne les fichiers de données à préserver
- A la BNF, le fichier METS est traité comme un fichier de métadonnées (le SIP est au format METS) ; il doit être conforme aux recommandations (profil METS) énoncées par la BNF, quant à sa structuration et à son contenu (valeurs) ; il est contrôlé et validé en entrée ; sa production est obligatoire

5.1.4 Constats

Les établissements disposent tous d'un inventaire des contenus numérisés à préserver dans un format structuré et, pour 5 établissements sur 6, dans un format standardisé (Dublin Core).

Les métadonnées de préservation, et notamment les métadonnées de provenance qui permettent de documenter le contexte dans lequel les fichiers image ont été produits, sont pauvres voire inexistantes.

Des incohérences dans la structuration des fichiers METS ont été relevées ; elles sont très vraisemblablement la conséquence d'un manque d'appropriation du format METS par les établissements.

Le format METS étant un format souple et générique, les établissements ont fait des choix d'implémentation qui ont pu être guidés soit par les recommandations et/ou spécifications des prestataires de préservation et notamment la BNF soit par la configuration des outils utilisés (OMEKA, NumaHOP). Ces choix s'avèrent donc différents selon les établissements et ont pu, pour un même établissement, évoluer au cours du temps.

La plateforme PAC du CINES est initialement basée sur Arcsys, mais en cours de bascule sur la solution VITAM. Les métadonnées descriptives et de préservation sont donc exprimées au format SEDA. Il conviendrait donc soit que le fichier METS soit converti au format SEDA, la solution de gestion du flux d'archivage choisie (NumaHOP par exemple) puisse directement produire un fichier au format SEDA.

5.2 MODALITES PROPOSEES

Les métadonnées rendent possibles les objectifs de la préservation numérique à long terme, qui sont de maintenir la disponibilité, la lisibilité, l'intelligibilité et l'authenticité des objets numériques. La compréhension des données, de leur utilisation, de leurs dépendances et de leur contexte permet en effet de les conserver de manière appropriée et de les documenter de manière suffisamment explicite pour que le contenu intellectuel puisse être conservé et compris à l'avenir.

L'objectif est double :

- Mieux documenter les collections des bibliothèques numériques par l'ajout de métadonnées supplémentaires,
- Normaliser la description des collections des bibliothèques numériques par la définition d'un modèle commun (profil METS) car l'hétérogénéité de l'implémentation du format METS est un frein à la mise en place d'une chaine de préservation unique et unifiée à l'échelle des établissements. Éventuellement, et en complément, établir le mapping (ou mise en correspondance) des balises METS avec les balises SEDA.

5.2.1 Enrichir la description des collections numérisées

Les métadonnées minimales à produire pour assurer la préservation des objets numériques sont les suivantes :

- Métadonnées d'identification : identifiants pérennes des ressources,

- Métadonnées descriptives (niveau document, article ou page) à exprimer en Dublin Core dans le fichier METS,
- Métadonnées techniques : métadonnées génériques (poids, format, empreinte numérique) ; métadonnées spécifiques pour les fichiers image (résolution, profil colorimétrique, profondeur d'encodage),
- Métadonnées de provenance : événement (numérisation, océrisation, création de la TdM), date, résultat, agent à exprimer selon le format PREMIS dans le fichier METS,
- Métadonnées de structure : liste des fichiers, ordre de lecture, rapports entre eux.

Un travail de recensement et d'identification de ces métadonnées reste à réaliser à l'échelle des établissements pour aboutir à la production d'un set de métadonnées commun à l'ensemble des établissements.

5.2.2 Normaliser la description des collections numérisées

NumaHOP est un logiciel Open source permettant de contrôler et de maîtriser la chaîne de numérisation, de l'import de notices bibliographiques à l'export de documents numérisés sur des bibliothèques numériques ou vers de l'archivage pérenne. Comme son utilisation se généralise au sein des membres de l'AMI, ceux-ci souaitent peut l'utiliser comme un levier afin de répondre à l'enjeu de convergence des pratiques documentaires de description des collections numérisées.

NumaHOP est composé de plusieurs modules fonctionnels permettant de :

- Convertir des notices au format UNIMARC ou EAD dans des formats interopérables : Dublin Core, Dublin Core qualifié,
- Réaliser des constats d'état pour les lots de documents à numériser envoyés vers les prestataires de numérisation,
- Recevoir les lots numérisés par le prestataire (images et métadonnées) et de les contrôler,
- Valider les unités documentaires numérisées (images + métadonnées) et de les exporter vers les diffuseurs et les archiveurs,
- Produire des fichiers OCR, METS.

Les établissements participants à l'AMI ont tous fait le choix d'utiliser le logiciel NumaHOP, ou envisagent de l'utiliser, pour traiter la chaine de numérisation. Le choix du même outil est un élément de convergence des pratiques documentaires.

Pour ces établissements, l'utilisation de NumaHOP a un coût : en effet, ils ont généralement opté pour une installation auprès d'un prestataire, en mode Saas (*software as a service*). L'installation de NumaHOP en interne est en théorie possible mais actuellement difficile : faciliter son déploiement est justement l'un des enjeux du projet CollEx-Persée « Ouvroir NumaHOP 2 », piloté par l'UAR Persée. Cependant, la convergence technique est insuffisante. Elle doit s'accompagner d'une réflexion collective – impliquant notamment CollEx-Persée, l'UAR Persée, l'ABES et le CINES – sur les métadonnées nécessaires pour la préservation.

Ces choix pourraient être consignés dans un profil METS, qui liste les différentes restrictions qu'un utilisateur ajoute aux contraintes existant déjà dans METS. Ce profil METS devrait, autant que possible, être compatible avec le profil METS défini par la BnF.

Par ailleurs, la documentation des choix dans un profil METS permettrait, dans un souci de traçabilité, de suivre l'évolution des règles de description des collections numérisées dans le temps et de contrôler, grâce à la mise en œuvre de schematrons, la conformité de l'implémentation des règles, en sortie de NumaHOP. À l'aune des pratiques des membres de l'AMI à la fin du GIS 1, NumaHOP apparaît comme une solution prometteuse pour produire des documents numérisés et des métadonnées permettant la préservation.

Enrichir et améliorer la qualité des fichiers METS produits par NumaHOP suppose au préalable de :

- Affiner les besoins des utilisateurs en matière de métadonnées de préservation,
- Définir, avec les établissements membres de l'AMI, un profil METS,
- Faire une analyse d'écart entre le profil souhaité et les fichiers METS actuellement produits par NumaHOP,
- Chiffrer le coût et planifier les évolutions.

Enfin, si on veut que les dépôts faits au CINES exploitent toutes les données enregistrées dans le fichier METS, il faudrait prévoir la possibilité d'exporter les mêmes métadonnées au format SEDA, alternativement ou en plus du format METS. L'autre solution serait de développer, en lien avec le CINES, un script de conversion du METS au SEDA.

Le financement de ces évolutions s'inscrit pleinement dans la politique nationale de numérisation portée par CollEx-Persée car il vient renforcer les logiques de mutualisation à la fois techniques et scientifiques et encourager le passage à l'échelle de la préservation des collections numérisées des bibliothèques. Il est le corollaire à l'ambition sous tendue par la politique nationale de numérisation.

Notons cependant que l'amélioration du fichier METS produit par NumaHOP n'est pas la seule évolution qu'il faudrait apporter, même s'il s'agit d'un point de départ. Une réflexion serait à mener sur le versement des données structurées et enrichies accompagnant les images numérisées ; et sur l'utilisation de NumaHOP pour des données nativement numériques. Parallèlement à cela un travail sur la correction des bugs encore existant est en cours. Il est porté par la communauté NumaHOP dans le cadre d'un groupe de travail sur la définition de la feuille de route de l'application. Mais ces réflexions dépassent le cadre de la présente étude.

Soulignons, enfin, que cet effort de construction d'une communauté de pratique doit s'accompagner d'un effort constant pour former et accompagner les producteurs de contenus aux bonnes pratiques de la préservation numérique, ce qui implique, au quotidien, une politique de gestion qualitative des données.

5.2.3 Préconisations

Réorienter les travaux du groupe de travail Préservation vers la production d'un set de métadonnées commun pour la préservation des collections numérisées

À partir de ce travail, et en lien avec tous les opérateurs concernés – UAR Persée, Abes, CINES –, définir un jeu de métadonnées nécessaires pour la préservation des collections numériques de l'ESR

Travailler avec la communauté des utilisateurs de NumaHOP pour enrichir le fichier METS produit par cette application, et faire financer les développements dans le cadre du dispositif CollEx-Persée, après une réflexion en coordination avec l'Abes et l'UAR Persée afin que le fichier produit soit conforme au profil METS souhaité.

Dans un souci d'interopérabilité, faire en sorte que NumaHOP puisse aussi produire un fichier de métadonnées au format SEDA en complément ou en remplacement du format METS, comprenant les mêmes métadonnées (par exemple en transformant, via une feuille de style, le fichier METS au format SEDA).

Dans un souci de traçabilité, promouvoir la mise en œuvre de PGD permettant de documenter les choix (et leur évolutions) opérés par les établissements et d'alimenter les métadonnées du METS.

En parallèle, étudier et évaluer les autres outils utilisés par les bibliothèques de l'ESR (chaîne Persée, collaborations avec Huma-Num) pour assurer la pérennité des données qu'elles produisent notamment lorsque la numérisation s'accompagnement d'enrichissements.

6. Annexes

6.1 Groupe de travail Preservation

Le groupe de travail Préservation a été constitué en septembre 2019 avec les principaux acteurs de la préservation numérique ainsi que des bibliothèques déjà impliquées sur le sujet : CINES, BnF, Huma-Num, SIAF, Abes, Persée, MESR, mais aussi Sciences Po Paris, La Contemporaine, le MNHN et la Bnu, ces deux derniers étant pilotes du groupe.

Pilotes du groupe de travail : Frédéric Blin (Bnu), Clément Oury (MNHN) et Catherine Désos-Warnier (chargée de mission GIS CollEx-Persée).

Membres à la création du groupe de travail : Viviane Boulétreau (Persée), Olivier Cian (ABES), Véronique De Kok (MESR), Louis Delespierre (MESR), Olesea Dubois (Sciences-Po), Yannick Grandcolas (BnF), Hélène Jouguet (HumaNum), Dominique Naud (SIAF), Stéphane Pouylleau (Huma-Num), Olivier Rouchon (CINES), Coline Silvestre (SIAF), Valérie Tesnière (La Contemporaine).

Etablissements membres de l'AMI Préservation : Bibliothèque interuniversitaire de la Sorbonne (Laurie Aooustet et Gilles Clément), BNU de Strasbourg (Rosanne Wingert), Bibliothèque Sainte Geneviève (Pauline Rivière), Bibliothèque de l'INHA (Pierre-Marie Bartoli), Bibliothèque de l'Institut français de Pondichéry, Bibliothèque de l'université Lyon 1 (François-Xavier Boffy), Bibliothèque du MNHN (Solenne Coutagne).

6.2 Convention de formation avec l'ENSSIB

enssib

DES / FTLV / INTRA / 22FC6

CONVENTION DE FORMATION PROFESSIONNELLE BIPARTITE [SUR-MESURE]

Entre les soussignés :

1-L'Ecole nationale supérieure des sciences de l'information et des bibliothèques,
Établissement public à caractère scientifique, culturel et professionnel
Situé(e) 17/21 bd du 11 Novembre 1918, 69623 Villeurbanne Cedex
Représenté(e) par sa Directrice, Madame Nathalie MARCEROU-RAMEL
N°SIRET : 196 924 591 00015 – Code APE : 8542Z – N° déclaration d'activité : 82 69 P003 269
Ci-après désigné(e) « **l'Enssib** »

Et

2 – La Bibliothèque nationale et universitaire de Strasbourg,
Établissement public national à caractère administratif, en tant qu'établissement support du GIS CollEx-Persée
Situé(e) au 5, rue du Maréchal Joffre, 67070 Strasbourg Cedex
Représenté(e) par son Directeur, Monsieur Alain COLAS,
N°SIRET : 180044067 00015 – Code APE : 91.01Z
Ci-après dénommée « **la Bnu** »,

Est conclue une convention de formation professionnelle en application de l'article L6353-4 du Code du Travail.

Article 1 : Objet du contrat

En exécution de la présente convention, l'Enssib s'engage à organiser l'action de formation intitulée « **Préserver les données numériques** » dans les conditions fixées par les 10 articles suivants.

Article 2 : Nature et caractéristiques de l'action de formation

2.1 - L'action de formation entre dans le champ d'application des dispositions relatives à la formation professionnelle continue prévue par l'article L6313-1 du Code du Travail.

2.2 – La formation est organisée en 2 temps :

- Un premier niveau de sensibilisation à distance : 2 webinaires de 3h, soit 6 heures de formation ;
- Un second niveau d'approfondissement en présentiel et à distance : 2 jours en présentiel à Paris (6h/jour) + un webinaire de 3h plusieurs semaines après, soit 12 heures en présentiel et 3 heures à distance.

2.3 – L'action de formation est réalisée selon un programme préétabli, qui en fonction d'objectifs déterminés, précise les moyens pédagogiques, techniques et d'encadrement mis en œuvre ainsi que les moyens permettant de suivre son exécution et d'en apprécier les résultats (cf. proposition de formation en annexe).

Objectifs de la formation

Objectifs du niveau 1 « sensibilisation » :

- Fournir un éclairage commun sur les pratiques et les enjeux de la préservation numérique
- Mettre à niveau les connaissances des participants dans le domaine de la préservation numérique

École nationale supérieure des sciences de l'information et des bibliothèques | *Membre associé de l'Université de Lyon*
17/21 bd du 11 novembre 1918 - 69623 Villeurbanne Cedex | SIRET : 196 924 591 00015 | +33 (0)4 72 44 43 43 | www.enssib.fr

Objectif du niveau 2 « approfondissement » :

> Dans le cadre de la préparation commune d'un groupement de commande, apporter les connaissances nécessaires à la préparation, la validation et la gestion d'une opération d'archivage pérenne.

Contenu de la formation (sous réserve de l'adaptation aux demandes spécifiques du groupe et des échanges avec le commanditaire)

Le niveau 1 abordera les thèmes suivants :

> Préparer et gérer une opération d'archivage : outils, méthodes, modèles conceptuels et formats, préparation des fichiers ;
> Assurer la sécurité des données ;
> Choisir le meilleur système de stockage des données ;
> Maîtriser les enjeux relatifs aux identifiants pérennes et aux métadonnées, notamment METS ;
> Partager une même méthode de gestion de projet.

Le niveau 2 abordera les thèmes suivants :

> Préparer et gérer une opération d'archivage : outils, méthodes, modèles conceptuels et formats, préparation des fichiers ;
> Assurer la sécurité des données ;
> Choisir le meilleur système de stockage des données ;
> Maîtriser les enjeux relatifs aux identifiants pérennes et aux métadonnées, notamment METS ;
> Partager une même méthode de gestion de projet.

Le niveau 2 doit être tourné vers la pratique (avec des études de cas), inclure de la manipulation de données et prendre en compte l'existant dans les établissements concernés.

Supports pédagogiques

L'Enssib proposera aux stagiaires d'accéder à sa plateforme pédagogique, afin qu'ils disposent, via un identifiant et un mot de passe, des supports pédagogiques fournis par le formateur en présentiel.

2.4 - Une évaluation sera proposée à la fin de la formation, sous la forme d'un questionnaire de satisfaction anonyme, fourni par l'Enssib, remis à chaque stagiaire afin qu'il puisse faire un retour plus personnel. Une synthèse de ces questionnaires sera remise au commanditaire.

2.5 - À l'issue de la formation, une attestation de suivi de formation, fournie par l'Enssib, sera délivrée aux stagiaires.

Article 3 : Organisation de l'action de formation

3.1 – La **responsabilité pédagogique** de l'action de formation est confiée à l'Enssib.
L'élaboration de son contenu est confiée à Marion MASSOL de l'AMUE.

3.2 – **L'effectif de la formation** est de :

> 25 stagiaires maximum pour le niveau 1, sans distinction de fonctions exercées ou de niveau hiérarchique (5 agents / établissement).

> 14 stagiaires maximum pour le niveau 2. Ces personnes devront obligatoirement avoir suivi le niveau 1 et être en situation de pilotage ou d'accompagnement direct du projet (2 agents / établissement).

3.3 – **Dates / horaires / lieu** :

Pour le niveau 1 : avril ou mai 2022 - Lieu : plateforme pédagogique de l'Enssib.

Pour le niveau 2 : mai, juin ou juillet 2022.

Les dates précises seront définies en concertation avec Marion MASSOL et devront être fixées au plus tôt.
Horaires indicatifs : 9h30 - 12h30 / 14h00 – 17h00 (avec une pause déjeuner).
Lieu : une salle de formation mise à disposition dans le centre de Paris + plateforme pédagogique.

3.4 – La salle de formation devra être équipée d'un ordinateur, d'un vidéoprojecteur, d'un tableau blanc (ou paperboard). Un pointeur serait un plus. Elle devra permettre le travail en groupe.

Article 4 : Obligations de l'Enssib

4.1 - L'Enssib s'engage à respecter le programme de l'action de formation.

4.2 – L'Enssib établit les attestations de suivi de formation sur la base des temps de connexion pour le distanciel et sur la base des feuilles de présence émargées par les stagiaires lors de chaque demi-journée pour le présentiel.

Article 5 : Obligations des pilotes du programme CollEx-Persée Préservation numérique

5.1 – S'engager à assurer la présence des participants conformément au programme prévu.

5.2 – S'engager à informer l'Enssib dans les 48 h. de tout incident ou évènement pouvant avoir des conséquences sur le déroulement de la formation. Les absences des stagiaires, qu'elles soient justifiées ou non, n'auront aucune incidence sur les dispositions financières prévues à l'article 6 de la présente convention.

5.3 – Pendant la durée de la formation, les stagiaires demeurent sous la responsabilité de leurs établissements d'origine pour tous les risques d'accident du travail et d'accident de trajet. L'obligation de déclaration incombe à chaque établissement d'origine. Les dommages éventuels que les stagiaires pourraient causer aux biens, personnels et usagers du lieu d'accueil du présentiel sont pris en charge par l'assurance responsabilité civile souscrite par leurs établissements d'origine respectifs.

Article 6 : Dispositions financières

6.1 – Le coût de la formation est évalué de la manière suivante :

Pour la **pédagogie** : **4 000€ TTC** (TVA non applicable) pour 2 webinaires de 3h chacun, 2 jours de formation en présentiel et un webinaire final de 3h et inclut :

- la préparation pédagogique de la formation ;
- la fourniture de supports de cours ;
- le suivi administratif ;
- tous les frais afférents aux formateurs ;
- l'accès à la plateforme pédagogique.

En sus, les **frais de de mission** (déplacement, hébergement et restauration) du ou des formateur(s) pourront être directement pris en charge par le commanditaire ou avancés par l'Enssib et refacturés ensuite au réel. Ajustables à la hausse ou à la baisse, en fonction du nombre d'intervenants qui seront finalement mobilisés, ils sont estimés à 1 000 euros environ pour la seule participation de Mme Massol.

Coût total de la formation : 5 000 € TTC (TVA non applicable).

6.2 – La prestation fera l'objet d'un bon de commande émis par la BNU de Strasbourg, dont le numéro devra impérativement être inscrit sur la facture.

La facturation s'effectuera via Chorus Portail Pro après la prestation, et sera payable par virement bancaire 30 jours à réception de la facture, aux coordonnées ci-dessous :

Code banque	Code guichet	N° compte	Clé	Domiciliation
10071	69000	0001004340	42	TP. LYON

IBAN (International Bank Account Number)	BIC (Bank Identifier Code)
FR76 1007 1690 0000 0010 0434 042	TRPUFRP1

Article 7 : Annulation ou interruption de la formation du fait de la Bnu

7.1 – En cas d'annulation au moins 15 jours francs avant le début de la formation, aucune somme n'est due par la Bnu.

7.2 – Si l'annulation intervient dans un délai inférieur à 15 jours francs avant le début de l'action de formation, il sera facturé à la Bnu, à titre d'indemnisation, 50% du coût de la formation, prévu à l'article 6 de la présente convention.

7.3 – En cas d'interruption par la Bnu en cours de formation, il lui sera facturé, à titre d'indemnisation, 50% du coût de la formation, prévu à l'article 6 de la présente convention.

Article 8 : Annulation ou interruption de la formation du fait de l'Enssib

8.1 – L'Enssib peut être amenée à annuler la session de formation. Dans cette hypothèse, l'Enssib s'efforcera de proposer à la Bnu une session de remplacement dans des termes à définir. Dans le cas où la mise en place d'une nouvelle session dans des conditions identiques s'avèrerait impossible, aucune somme ne peut être exigée de la Bnu.

8.2 – En cas d'interruption de la formation par l'Enssib, le montant dû par la Bnu sera celui correspondant aux heures de formation réellement effectuées.

Article 9 : Durée de la convention

La présente convention prend effet à compter du 1° avril pour s'achever le dernier jour de la formation. Elle n'est pas reconductible.

Article 10 : Modification de la convention

Toute modification de la présente convention doit être effectuée par voie d'avenant.

Article 11 : Litige

Tout litige qui pourrait intervenir dans le cadre de l'interprétation de la présente convention, sera au préalable réglé à l'amiable par les signataires. En cas de désaccord persistant, les co-contractants s'en remettront à la juridiction compétente, le tribunal administratif de Lyon.

Fait en deux exemplaires originaux.

À Villeurbanne, le	À Strasbourg, le
La Directrice de l'Enssib	**Le Directeur de la Bnu et du GIS CollEx-Persée**
[Cachet et signature]	[Cachet et signature]
ENSSIB Villeurbanne	
Nathalie MARCEROU-RAMEL	**Alain COLAS**

6.3 LETTRE DE MISSION

Appel d'offre pour l'accompagnement pour la constitution d'un groupement de commandes en vue d'une prestation de préservation numérique

Appel d'offre

N°2022 - 04

Objet :	Accompagnement pour la constitution d'un groupement de commandes en vue d'une prestation de préservation numérique
Établissement :	Bibliothèque nationale et universitaire de Strasbourg 5, rue du Maréchal Joffre B. P. 51029 67070 Strasbourg Cedex
Date limite de réponse :	15 septembre 2022
Type de marché :	Prestation de service

CONTEXTE DE LA MISSION

Le GIS CollEx-PERSEE (https://www.collexpersee.eu/), par l'intermédiaire administratif de la Bibliothèque nationale et universitaire de Strasbourg, ouvre une offre de mission pour un accompagnement visant la constitution d'un groupement de commandes pour une prestation de préservation numérique.

L'objectif final de la mission est l'attribution à un prestataire de préservation numérique, d'un marché issu d'un groupement de commandes réunissant les établissements suivants :

- Bibliothèque nationale et universitaire de Strasbourg
- Bibliothèque du Muséum national d'histoire naturelle de Paris
- Bibliothèque interuniversitaire de la Sorbonne
- Bibliothèque de l'Institut national d'histoire de l'art
- Bibliothèque Sainte-Geneviève
- Bibliothèque de l'Université Lyon I
- Bibliothèque de l'Institut français de Pondichéry

Il est à noter que des établissements pourront être amenés à se retirer du dispositif, notamment s'ils ne sont pas dans la capacité technique ou financière de se conformer aux conditions d'appartenance qui seront collectivement élaborées.

OBJET DE LA MISSION

La mission a pour objet un service de préservation des données numériques fournies par les établissements membres du groupement de commandes. Ces données numériques sont issues d'opérations de numérisation portant essentiellement sur des documents de bibliothèque (imprimés, archives et manuscrits, iconographie…). Un des principaux objectifs de l'opération est que le volume global représenté par les données issues des différents établissements constitue un ensemble homogène et d'ampleur suffisante pour pouvoir bénéficier d'offres commerciales intéressantes sur le plan financier de la part de prestataires de préservation numérique, que ceux-ci soient des établissements publics ou des acteurs privés.

CALENDRIER DE LA MISSION

La mission est établie pour la durée suivante : 15 octobre 2022 au 30 juin 2023.

MONTANT DE LA MISSION

Le montant de la mission ne pourra pas dépasser 10 000€ HT.

DESCRIPTION DE LA MISSION

Le prestataire aura donc les tâches suivantes :

- Analyser les données numériques proposées par chaque établissement
- Conseiller les établissements sur la structuration de leurs données, dans l'optique de parvenir à des données répondant aux mêmes caractéristiques dans chaque établissement
- Rédiger le cahier des clauses techniques particulières du marché
- Participer en tant qu'expert à l'analyse des offres reçues

CONFIDENTIALITÉ

Le prestataire s'engage à ne pas divulguer les informations et documents de quelque nature que ce soit, relatifs au fonctionnement de la Bnu, qu'il aurait pu recueillir, obtenir ou dont il aurait pu avoir connaissance au cours de l'exécution du présent marché.

De même, la Bnu s'engage à garder strictement confidentiel l'ensemble des informations échangées, notamment les renseignements techniques et commerciaux.

6.4 CARACTERISTIQUES DES COLLECTIONS NUMERISEES : ETAT DES LIEUX

Vue d'ensemble des fonds numérisés suite aux entretiens avec les établissements

Etablissement	Début des opérations de numérisation	Nature des fonds	Conditions de numérisation (interne / externe)	Formats des fichiers	Formats des métadonnées descriptives	Environnement de conservation des fichiers
Bibliothèque nationale et universitaire de Strasbourg	2000	Documents patrimoniaux : manuscrits, documents iconographiques, documents cartographiques, objets, texte imprimé, papyrus Documents nativement numériques : recensement en cours	Numérisation en interne (documents les plus précieux) + en externe	TIFF JPG (format de diffusion)	Dublin Core	SPAR (pour une partie des corpus, notamment la presse numérisée)
Bibliothèque du Muséum national d'histoire naturelle de Paris	2000	Monographies, périodiques, archives, manuscrits, cartes et plans, iconographies... (Archives nativement numériques à venir dans un futur incertain)	Numérisation en interne (1/3) et en externe (2/3)	TIFF JPG (format de diffusion)	Excel qui comprend de l'EAD et de l'UNIMARC	Bandes + serveurs
Bibliothèque interuniversitaire de la Sorbonne	Début des années 2000	Fonds de la Réserve/documents patrimoniaux numérisés : Manuscrits, archives, images, documents cartographiques, monographies Fonds nativement numériques / documents sonores et audiovisuels : archives orales, podcast, vidéos	Numérisation en interne (début années 2000)+ externe pour certains corpus (depuis 2019)	TIFF JPG + PDF (format de diffusion) MP3, MP4, WAVE (archives nativement numériques)	DublinCore Métadonnées non standardisées pour les archives sur la chaine YouTube	Serveurs + ce qui est sur la bibliothèque numérique est sauvegardé à la DSI sur bandes
Bibliothèque de l'Institut national d'histoire de l'art	2002	Archives, manuscrits, dessins, estampes, monographies imprimées, photographies. Numérisation des microfilms jusqu'au début des années 2010, privilégie depuis la numérisation des originaux	Prestations externes de numérisation Numérisation interne à la demande par le "Service photo"	TIFF (master) JPG (format de diffusion) PDF multicouche (format de diffusion)	Dublin Core XML-ALTO	Serveurs redondés, contenant le lecteur réseau (Numerisation (N:)) pour ce qui n'est pas publié en ligne
Bibliothèque Sainte-Geneviève	2011	Documents patrimoniaux : Manuscrits, archives, imprimés (périodiques, monographies, incunables, partitions etc.), estampes, cartes et plans.	Prestations externes de numérisation	PNG TIFF (fichiers ont été convertis au format PNG) JPG2000 (depuis 2021)	Dublin Core XML-ALTO TEI	CINES depuis 2011
Bibliothèque de l'Université Lyon 1	2009	Fonds anciens : monographies imprimées, périodiques Thèses et mémoires nativement numériques	Prestations externes de numérisation	TIFF PDF (format de diffusion) JPG (format de diffusion)	Dublin Core	CINES pour les thèses BNF (périodiques médicaux + Thèses pour le XIXe siècle)

Vue d’ensemble des métadonnées :

Etablissement	Format des métadonnées	Formats des fichiers de métadonnées	Métadonnées de gestion	Contexte de production des fichiers de métadonnées	Remarques
BNU	Dublin Core EAD Unimarc METS	Les métadonnées au format Dublin Core sont encapsulées dans le fichier METS (section dmdSec-informations bibliographiques) Les métadonnées au format EAD sont exprimées dans un document XML Les métadonnées au format UNIMARC sont exprimées dans un document .mrk Les métadonnées au format METS sont exprimées dans un document XML	Fichier METS normalisé selon les recommandations de la BNF	Interface de catalogage Calames pour le fichier XML-EAD Cataloge Sudoc pour le fichier .mrk SPAR pour le fichier METS	Champs Dublin Core : Title, Creator, Date, Publisher, Contributor, Type, Format, Language, Coverage spatial, Coverage temporal, Rights
MNHN	Unimarc Grille d'indexation interne	Les métadonnées au format UNIMARC sont exprimées dans un document .mrk Les métadonnées de la grille d'indexation interne sont exprimées dans un document Excel (-> leur rôle est de préparer l'intégration dans le portail documentaire)	–	Cataloge Sudoc pour le fichier .mrk Prestataire de numérisation pour le fichier .xls	Métadonnées renseignées dans la grille d'indexation : Type, Identifiant_MARC, Date_New_Item, Identifiant, Cote_VAL, Type_Doc, Droits_Mention, Watermark, Repertoire_Racine, Repertoire_Serie, Nom_du_Fichier, Elt_Num, Tuilage, Contrat, OAI_SET
BIS	Dublin Core METS	Les métadonnées au format Dublin Core sont encapsulées dans le fichier METS (section dmdSec-informations bibliographiques) + lien vers la notice Sudoc dans le fichier METS Les métadonnées au format METS sont exprimées dans un document XML	Fichier METS produit par OMEKA-Classic	Omeka-classic pour le fichier METS (plug-in MetsExport)	Présence de fichiers METS que sur une partie du corpus (corpus traité via Omeka-Classic -> jusqu'en 2020) Champs Dublin Core (non systématiquement tous renseignés) : Creator, Title, Coverage, Format, Type, Identifier Source, Date, Description, Contributor, Publisher, Relation, Subject, Language
INHA	Dublin Core ALTO Grille d'indexation des TDM EXIF	Les métadonnées au format ALTO sont exprimées dans un document XML Les métadonnées de la grille d'indexation de la table des matières sont exprimées dans des documents Excel, Txt et Pdf Les métadonnées au format Dublin Core sont exprimées dans un document Excel	Métadonnées EXIF encapsulées dans les propriétés des fichiers TIFF ou PNG	Appareil de prise de vue pour les métadonnées au format EXIF Processus d'océrisation pour les métadonnées au format ALTO	Les notices descriptives en Dublin Core sont plus ou moins détaillées, selon la nature des documents
BSG	Dublin Core METS sip.xml	Les métadonnées au format Dublin Core sont encapsulées dans le fichier METS (section dmdSec-informations bibliographiques) Les métadonnées au format METS sont exprimées dans un document XML Les métadonnées correspondant à la description des objets à verser dans la plateforme PAC du CINES sont exprimées dans un document XML -> format sip.xml de description des métadonnées de la plateforme d'archivage PAC du CINES Les métadonnées de la grille d'indexation de la table des matières sont exprimées dans des documents Txt	Fichier METS produit par les prestataires de numérisation puis Numa-HOP	Prestataires puis Numa-Hop pour le fichier METS et le fichier sip.xml Prestataires de numérisation pour le fichier METS	Conversion en cours du format sip.xml vers le format SEDA 2.1 Variation dans la construction des fichiers METS, en fonction des prestataires
Lyon 1	Dublin Core METS sip.xml	Les métadonnées au format Dublin Core sont encapsulées dans le fichier METS (section dmdSec-informations bibliographiques) Les métadonnées correspondant à la description des objets à verser dans la plateforme PAC du CINES sont exprimées dans un document XML -> format sip.xml de description des métadonnées de la plateforme d'archivage PAC du CINES	Fichier METS produit par Numa-HOP	Numa-Hop pour le fichier METS et le fichier sip.xml	Les fichiers METS et sip.xml n'existent que pour les thèses préservées au CINES Conversion en cours du format sip.xml vers le format SEDA 2.1