

Groupe de travail interministériel

Étude de faisabilité
d'un outil national de production
en EAD

Rapport final

Novembre 2015

Sommaire

Avant-propos.....	4
Résumé exécutif.....	5
I. Contexte de l'étude, méthodologie.....	7
1. Contexte général.....	7
Historique de l'EAD en bibliothèque.....	7
Contexte actuel.....	7
2. Lettre de mission confiée au groupe de travail.....	9
3. Méthodologie mise en place.....	10
Analyse des outils existants.....	10
Définition des besoins fonctionnels.....	10
Elaboration d'une typologie des usagers de l'outil et adéquation avec les besoins fonctionnels.....	11
Conception d'un dispositif par marches progressives, déclinées en 3 scénarios.....	11
Examen de solutions impliquant l'éditeur XXE.....	12
Rédaction du rapport.....	12
4. Périmètre scientifique et fonctionnel de l'outil de production national.....	13
II. Besoins fonctionnels identifiés et modalités de mise en œuvre.....	14
1. Gestion des autorités	15
2. Gestion des contrôles	15
3. Gestion des droits	16
4. Convivialité, facilité d'installation et maintenance de l'outil	17
5. Conversion des données en amont et en aval de l'outil	17
6. Gestion des unités physiques.....	18
7. Formation initiale/assistance/boîte à outils	19
Questions de gouvernance et d'organisation.....	19
III. Scénarisations envisagées.....	21
1. Agencement des besoins.....	21
2. Solutions écartées par le groupe de travail.....	22
3. Examen de l'éditeur XXE.....	23

4. Scénarios retenus par le groupe de travail.....	24
5. Coûts comparés.....	28
6. Conséquences stratégiques et questions à résoudre.....	29
Conclusion.....	31
Annexes.....	32
Annexe 1 – Déclaration d’intention du 17/06/2014 « Concertation nationale autour d'un outil commun de catalogage en EAD ».....	33
Annexe 2 – Composition du groupe de travail « Outil national EAD ».....	36
Annexe 3 – Liste des réunions.....	37
Annexe 4 – Sept années d’encodage dans Calames Evolutions d’un réseau de production en EAD (2008-2014).....	38
Annexe 5 – Scénariser un projet d’outil national EAD : le point de vue de l’équipe Calames	42
Annexe 6 – Solutions présentées au groupe de travail par la BnF-CCFr le 16/09/2015.....	45

Avant-propos

Le présent rapport est le fruit d'une année de travail et d'échanges au sein d'un groupe constitué de représentants de la Bibliothèque nationale de France, de bibliothèques municipales et de bibliothèques de l'enseignement supérieur, de l'Equipex Biblissima et du Service du livre et de la lecture du Ministère de la culture et de la communication¹. Il repose sur un certain nombre de constats, de partage d'expériences et surtout de compétences techniques avérées des participants.

Il vise avant tout à dresser un état des lieux des outils existants et à faire un certain nombre de propositions de scénarios pour un outil national de production en EAD destiné avant tout aux bibliothèques.

Ce rapport ne saurait donc régler un certain nombre de questions pendantes, notamment sur les plans politique et stratégique de l'outil national, dont le comité de pilotage et les tutelles devront se saisir.

¹ Cf. annexe 2.

Résumé exécutif

Le développement de l'EAD en bibliothèque est aujourd'hui à un carrefour. Après la finalisation du guide des bonnes pratiques de l'EAD dans les bibliothèques françaises, les établissements et les réseaux sont soit en attente d'un outil de catalogage leur permettant de les mettre en œuvre, soit dans le besoin de moderniser les leurs.

Le rapport du groupe de travail, s'inscrivant dans le cadre de la mission qui lui a été confiée en juin 2014 et tenant compte de l'évolution du contexte et des choix opérés par des réseaux membres depuis cette date, parvient aux conclusions suivantes :

1- **Un outil national de production en EAD demeure une nécessité** pour les établissements et les réseaux pour des raisons économiques et scientifiques. Il doit **être appuyé sur un référentiel national d'autorités**, dont la réalisation et le fait qu'il soit partagé par des outils de catalogage en EAD qui pourraient rester différenciés dans un premier temps, constitueraient un pas important vers la mutualisation de la production de données.

2- Le groupe de travail met également en lumière **six autres fonctionnalités** importantes pour un outil de production national en EAD : gestion des contrôles, gestion des droits, ergonomie d'utilisation et de maintenance de l'outil, conversion de données en amont et en aval de l'outil, gestion des unités physiques, boîte à outil et maintenance scientifique.

3- Il a envisagé le développement de l'ensemble de ces fonctionnalités dans le cadre de **marches successives**. La première serait de doter les établissements qui en sont dépourvus d'un outil de production dont la réalisation serait une première étape vers la mise en place d'un outil national EAD. Cela concerne essentiellement le réseau CCFR. La seconde marche serait la mise en cohérence de l'ensemble des producteurs de données en EAD par l'utilisation d'un référentiel commun d'autorités et d'un système cohérent d'identifiants pérennes. La troisième marche serait la mise en œuvre d'un outil national de production commun aux établissements et aux réseaux volontaires.

4- Pour la réalisation de ces différentes marches, le groupe de travail a envisagé **trois pistes** : le développement d'un outil *ex nihilo*, qui est un scénario détaillé dans le but de servir d'étalon pour le travail qui pourrait être réalisé à partir des outils existant, à savoir un développement de PIXML, l'outil EAD de la BnF (deuxième piste) ou un développement de Calames, l'outil EAD de l'ABES (troisième piste).

5- **Dans l'état actuel des réflexions et des projets des différents réseaux et institutions**, le groupe de travail préconise que :

a/ Les choix faits par le réseau CCFR s'inscrivent comme une étape vers la réalisation d'un outil partagé par d'autres réseaux et établissements.

b/ Les réflexions menées par la BnF et l'ABES sur le remplacement de l'éditeur XML soient mutualisées et débouchent sur des choix concertés et communs.

c/ Le projet de mutualisation de la production de données d'autorité au niveau national se concrétise dans les meilleurs délais en respectant les étapes nécessaires pour y parvenir.

d/ Le comité de pilotage et les tutelles ministérielles envisagent avec la BnF et l'ABES les moyens nécessaires pour mettre en œuvre les propositions faites par ces établissements.

e/ Un suivi et un accompagnement des étapes de la mise en œuvre d'un outil national partagé puisse être assuré par les différents acteurs du projet (BnF, ABES, établissements concernés, etc.).

I. Contexte de l'étude, méthodologie

1. Contexte général

Historique de l'EAD en bibliothèque

Une première étape significative dans le signalement en ligne des collections de manuscrits des bibliothèques publiques françaises (bibliothèques municipales et bibliothèques universitaires) a consisté, dans les années 2003-2008, en une informatisation du *Catalogue général des manuscrits des bibliothèques publiques de France* (CGM).

Ces données structurées au format EAD (Encoded archival description) sont désormais accessibles dans la base Manuscrits du Catalogue collectif de France (CCFr)², ainsi que dans le Catalogue en ligne des archives et des manuscrits de l'enseignement supérieur (Calames)³. Parallèlement à l'encodage du CGM, les bibliothèques universitaires se sont en effet dotées d'un outil de catalogage spécifique, Calames (2006-2007)⁴, qui permet à la fois la mise à jour des données du CGM et aussi la production de notices *ex nihilo*. Cet outil est géré par l'ABES.

Dans le même temps, la BnF a elle aussi procédé à l'informatisation de ses catalogues de manuscrits. Elle s'est également dotée d'un outil de catalogage et de gestion des instruments de recherche en EAD, PiXML, et d'une interface de consultation : BnF Archives et manuscrits (BAM).

La mise en place d'un groupe de travail, rassemblant depuis juin 2008 des représentants des bibliothèques relevant du MCC et du MESR, a abouti en 2013 à la publication d'un *Guide des bonnes pratiques de l'EAD en bibliothèque*⁵, qui a permis de définir un usage commun de l'EAD dans les bibliothèques françaises⁶. Ce guide constitue de fait la première pierre à partir de laquelle un outil de production national en EAD peut être mis en œuvre.

Contexte actuel

Actuellement, un certain nombre de bibliothèques ne dispose pas d'un outil de production et de correction de notices descriptives en EAD. C'est notamment le cas des bibliothèques du réseau du CCFr : les établissements qui souhaitent mettre à jour leur catalogue de manuscrits

² La base Manuscrits du CCFr compte 470 établissements placés majoritairement sous la tutelle du ministère de la Culture.

³ Les données de la Bibliothèque de l'Arsenal (Paris) se trouvent, par exception, dans le catalogue BnF Archives et Manuscrits.

⁴ 50 établissements déployés en 2015. <http://www.abes.fr/Calames/Le-reseau-Calames/Les-etablissements-membres-du-reseau-Calames>, consulté le 13/07/2015.

⁵ <http://bonnespratiques-ead.net/>, consulté le 13/07/2015.

⁶ L'EAD en bibliothèques se distingue des pratiques en services d'archives par une attention particulière envers toutes les formes d'indexation documentaire, et par la prise en compte d'une diversité documentaire, moins sérielle qu'en archives et notamment exprimée par la recommandation DeMArch. Pour autant, les appropriations de l'EAD par archivistes et bibliothécaires ne s'excluent pas et il apparaît nécessaire (et la solution de production projetée par le GIT doit permettre) d'envisager l'avenir de l'EAD en France à l'échelle de ces deux communautés, indépendamment des différences de choix d'outils et de modes d'encodage.

travaillent sur des éditeurs XML installés en local et les mises à jour du CGM se font encore par chargement manuel d'instruments de recherches dans la base Manuscrits du CCFr alors que le volume de données contenu dans la base CGM, l'ampleur des collections restant à signaler dans les BM et la dynamique enclenchée au cours des dernières années justifient pleinement la mise à disposition d'une application de catalogage et de versement automatique des instruments de recherche dans le CCFr. **La mise en œuvre d'un outil de production en EAD à destination des bibliothèques du réseau CCFr a été déclarée prioritaire et annoncée lors des journées des pôles associés (2-3 octobre 2014).**

Par ailleurs, les deux catalogues Calames et BnF Archives et manuscrits sont en cours ou doivent prochainement faire l'objet d'évolutions importantes, actées ou prévisibles.

Calames a fait l'objet d'une étude collective faisant ressortir un satisfecit global sur le fonctionnement de l'outil, réserves faites de certaines lenteurs et inaccessibilités fonctionnelles liées à l'outil XMetal⁷. Le développement du périmètre scientifique de Calames (manuscrits, puis archives, puis fonds y compris iconographiques ou audiovisuels) comme des besoins structurels constatés (universités se dotant de services d'archives, prise en charge par certains services communs de la documentation de la mission de gestion des archives institutionnelles, et plus généralement, rapprochement des archives institutionnelles et des archives de la recherche) justifient les moyens à consacrer à cet outil côté MESR.

Pour la BnF, après un rapport sur l'avenir de l'EAD au sein de l'établissement (2013), une refonte de l'outil de production et de gestion (PIXML) et une refonte de l'outil de consultation (BAM) sont en cours pour un achèvement programmé courant 2016. Le projet de refonte a été résolument inscrit par la BnF dans un contexte de rapprochement avec le MESR : la convention BnF/ABES en fait état et le projet est présenté comme suit dans le contrat de performance 2014 :

La BnF s'engage au cours de la période à développer un outil de production en EAD (format permettant de structurer les descriptions de manuscrits et de documents d'archives), en remplacement de l'actuel catalogue BAM (BnF Archives et manuscrits). Il s'agit d'acter la généralisation de l'EAD à un périmètre documentaire élargi et cohérent et de doter les personnels d'un outil de travail répondant aux fonctionnalités requises. La BnF s'attachera à la mutualisation avec les autres institutions concernées, dans l'optique d'un futur outil national de production en EAD qui constituerait la brique logicielle indispensable à l'enrichissement et au développement du catalogue national des manuscrits coordonné par le CCFr. (...) Sera conduit en parallèle un chantier de refonte de l'interface publique du catalogue BnF-Archives et manuscrits en cohérence avec le chantier de refonte de son outil de production. La convergence des évolutions du CCFr avec celles des autres catalogues sera également privilégiée⁸.

D'autre part, s'insère dans ce panorama un nouvel acteur : l'équipex Bibliissima. Porté par le Campus Condorcet, cette structure se définit comme « un observatoire du patrimoine écrit du Moyen Âge et de la Renaissance » ; outre les partenaires du Campus Condorcet, elle rassemble la BnF, le Centre d'études supérieures de la Renaissance (CNRS/université de

⁷ Les principales conclusions de cette étude ont été présentées lors de la journée des correspondants Calames du 27 mai 2014 : <https://calames.wordpress.com/2014/06/06/retour-sur-la-journee-du-reseau-calames/>, consulté le 13/07/2015.

⁸ http://www.bnf.fr/documents/contrat_performances_2014.pdf, p. 9, consulté le 09/12/2014.

Tours), le laboratoire CIHAM (Histoire, Archéologie, Littératures des mondes chrétiens et musulmans médiévaux : CNRS - Université Lyon 2 - EHESS - ENS de Lyon - Université d'Avignon et des Pays de Vaucluse - Université Lyon 3), l'École nationale des chartes, l'IRHT, la MSH de Caen, le laboratoire SAPRAT (EPHE) et le CRAHAM (CNRS/université de Caen). Plus précisément,

Bibliissima est une bibliothèque numérique sur le web, donnant accès de façon simple et coordonnée à un ensemble massif et complexe de documentation sur les manuscrits et les imprimés anciens, les textes qu'ils transmettent, leur circulation et leurs lecteurs, du VIII^e au XVIII^e siècle (documents anciens numérisés, bases de données documentaires, éditions de textes, outils pour la compréhension de ces documents et la production de données nouvelles). C'est un équipement de données pour la recherche, la formation et la diffusion visant à la fois les chercheurs et un public plus large. Sa réalisation s'articule en deux volets : l'observatoire (les outils) ; les corpus de données scientifiques (les contenus)⁹.

Bibliissima a d'ores et déjà acquis les sources du logiciel XML Mind, assorti d'une licence perpétuelle d'utilisation au sein des neuf partenaires du projet et dans le cadre des missions du consortium (utilisation de l'EAD, de la TEI, etc.).

Enfin, **le rapprochement opéré entre la BnF et l'ABES pour favoriser l'évolution des catalogues nationaux vers le web des données liées** a abouti en novembre 2014 à la promulgation de préconisations communes¹⁰. Ce partage de vues au niveau national dessine un contexte favorable à la création d'un outil de production commun en EAD, qui vise à renforcer la normalisation des données.

Forts de ce constat, le MCC-SLL, le MESR-DISTRD, la BnF, l'ABES et Bibliissima se sont mis d'accord lors d'une réunion commune le 17 juin 2014 sur la pertinence de réfléchir à un outil commun de production. Pour ce faire, une étude de faisabilité a été confiée par le comité de pilotage à un groupe de travail qui s'est réuni entre septembre 2014 et octobre 2015.

2. Lettre de mission confiée au groupe de travail

La déclaration d'intention¹¹, validée conjointement par le MCC, le MESR, la BnF, l'ABES et l'équipex Bibliissima, suite à la réunion du 17 juin 2014, a défini les missions attribuées au GIT Outil de production national en EAD, à savoir :

- Faire un état des lieux des solutions techniques actuelles pour le catalogage en EAD,
- Définir un noyau consensuel de besoins fonctionnels attendus d'un outil commun de catalogage en EAD : le groupe se concentrera sur les questions logicielles pour la description en EAD (et accessoirement en EAC), en prenant en compte l'évolution probable des schémas et la nécessité de pouvoir gérer simultanément les formats EAD 2002 et EAD 3. Le groupe réfléchira également à l'opportunité d'inclure des fonctionnalités de gestion des unités physiques et de communication des documents,

⁹ <http://www.bibliissima-condorcet.fr/fr>, consulté le 09/12/2014.

¹⁰ http://www.bnf.fr/documents/1411_transition_bibliographique.pdf, consulté le 13/07/2015.

¹¹ Cf. annexe 1.

- Élaborer des scénarios et options détaillant les principales spécifications du futur outil. Présenter la dimension économique de chaque scénario : ressources humaines, budget de développement, répartition des charges, etc.,
- Produire un rapport à l'issue de l'étude de faisabilité.

La réunion du comité de pilotage du 19 janvier 2015 a confirmé ces attendus et permis notamment d'affirmer la nécessité d'adosser cet outil de production à un référentiel national d'autorités et de préciser les attentes du COPIL en matière de scénarisation des besoins, qui doit s'accompagner d'un chiffrage des coûts attendus.

3. Méthodologie mise en place

Analyse des outils existants

S'appuyant sur une grille d'analyse antérieure des outils de production et de publication en XML-EAD et EAC, constituée par le MCC/SLL, le GIT s'est livré à un audit des produits actuels disponibles, via

- une exploration directe de ces outils (ex. Yale/Xmetal, Archivist Toolkit) ou via des présentations (ex. logiciel CD-EAD de la société CD-Script, déployé à la BM de Roubaix)
- des démonstrations par les utilisateurs membres du groupe de travail : PiXML (BnF) ; Mnesys archives (version déployée à la BDIC) ; XML Mind – XML Editor (MSH de Caen)

Le dispositif d'analyse a été complété par l'audition de Pascal Tanesie (Archives nationales), venu présenter le Système de gestion d'archives (SIA) des Archives nationales. Le Service interministériel des Archives de France, représenté par Claire Sibille et Alice Motte, a enfin fait profiter le groupe d'un panorama des logiciels utilisés dans les services territoriaux d'archives, au sein duquel seuls deux outils s'avèrent permettre la production en XML-EAD natif : Ligeo archives et Mnesys archives. L'attention du GIT a également été attirée à cette occasion sur le logiciel libre ATOM, développé avec l'appui du Conseil international des archives.

Le GIT regrette toutefois que l'étude comparative des éditeurs XML conduite par la Direction des systèmes d'information de la BnF soit lancée en parallèle de cette mission et donc non disponible au moment de la publication du présent rapport.

Définition des besoins fonctionnels

À partir des pratiques des membres du groupe de travail, des remontées émanant des différents réseaux et de l'analyse des outils existants, le groupe de travail s'est attaché à définir **un socle commun de besoins fonctionnels**, prenant en compte l'évolution des usages mais aussi des techniques et des normes de description (notamment le passage de l'EAD 2002 à l'EAD3). Le respect du *Guide des bonnes pratiques de l'EAD en bibliothèque* a été considéré comme un préalable absolu à toute fonctionnalité.

Elaboration d'une typologie des usagers de l'outil et adéquation avec les besoins fonctionnels

Huit catégories d'utilisateurs ont été identifiées, ayant des besoins de recours à cet outil plus ou moins ponctuels :

- a. Départements de la BnF,
- b. Grands établissements et bibliothèques de l'ESR faisant partie du premier cercle Calames,
- c. BMC de grandes villes : collections importantes et moyens humains de traitement de celles-ci,
- d. BMC de villes moyennes ou BM possédant des fonds remarquables mais des moyens souvent limités,
- e. SCD ou/et services d'archives des universités,
- f. BM sans fonds remarquables : peu de documents à décrire – peu de moyens et/ou de compétences en EAD,
- g. Laboratoires, UMR, USR ne versant pas aux Archives nationales,
- h. Structures associatives et sociétés savantes.

Chacune des fonctionnalités définies par le GIT a été systématiquement revue en fonction de ces catégories d'établissement pour vérifier l'adéquation entre le besoin et l'offre. **Chacun des scénarios proposés au sein de ce rapport entend donc correspondre, aussi bien en termes d'ergonomie que de formation ou de facilité d'installation, à l'ensemble des établissements appartenant à cette typologie** : dans le cadre d'un outil national de production, il n'a pas paru pertinent de proposer des déclinaisons fonctionnelles en fonction de la typologie des usagers.

Conception d'un dispositif par marches progressives, déclinées en 3 scénarios

Pour le groupe de travail, l'objectif final serait de mettre en place à la fois un outil de production *stricto sensu*, mais aussi

- 1) d'asseoir cet outil sur une base unique d'autorités, permettant à chacun des catalogueurs non seulement de profiter des deux grands réservoirs nationaux existant BnF et IdRef mais aussi de les enrichir soit par lots (ex. fichiers d'autorités locaux de la BM de Lyon), soit ponctuellement ;
- 2) d'appuyer cet outil sur une infrastructure chargée à la fois de l'hébergement des données, de la maintenance corrective et applicative et de l'accompagnement des usagers (formation, hotline, etc.).

Au vu des calendriers contraints de certains des partenaires (le besoin exprimé par le CCFr est la mise à disposition d'un outil à la fin de l'année 2016 pour répondre au contrat de performance de la BnF), le groupe de travail a donc **réfléchi à une scénarisation « progressive », correspondant à plusieurs étapes de développement de l'outil qualifiées de « marches »**, allant de la mise à disposition d'un éditeur XML simple et convivial à la mise en place d'une interface unique et partagée de production.

Ce dispositif a ensuite été **décliné en trois scénarios, s'appuyant soit sur la récupération de briques fonctionnelles existantes issues de Calames et Pixml, soit sur une création *ex nihilo***. Ces trois scénarios ont fait l'objet d'un chiffrage, exprimé en coût de développement,

de formation et de maintenance, pour lesquels les équipes techniques de la BnF et de l'Abes ont été sollicitées.

Examen de solutions impliquant l'éditeur XXE

Enfin, en vue d'asseoir ou d'invalider certains aspects techniques dans l'exposition des scénarios, des échanges et tests portant sur XXE et sur le récent développement d'un plugin collaboratif réalisé dans le cadre de Biblissima ont été menés au sein de la BnF et de l'ABES en septembre-octobre 2015. Cet examen ne peut tenir lieu d'une véritable évaluation technique des solutions d'édition XML, non plus que d'une préfiguration ou d'un protoypage que ni la lettre de mission du GIT ni son calendrier ne permettaient.

Rédaction du rapport

La rédaction du rapport a enfin été partagée entre les différents membres du GIT et fait l'objet d'une relecture collective lors de la dernière réunion du groupe de travail.

4. Périmètre scientifique et fonctionnel de l'outil de production national.

Il importe enfin de rappeler en exergue certains points de définition et/ou de périmètre couvert par cet outil national :

- L'outil national de production en EAD est ouvert au signalement de tous types d'objets, dans la mesure où ils se prêtent à la description en EAD : cela implique évidemment les archives papier et les manuscrits, mais aussi les archives audiovisuelles, les archives numériques, les documents iconographiques ou les objets, et plus largement tous documents décrits dans une logique arborescente et contextuelle.
- L'outil national de production en EAD est ouvert à tous types d'institutions publiques – quel que soit leur rattachement institutionnel – ou d'autres acteurs relevant du droit public ou du droit privé (ex associations ...), dès lors que l'accès aux données est libre, gratuit et possible via le CCFr. La politique tarifaire de mise à disposition de cet outil est cependant à définir.
- L'outil national de production en EAD reste un outil professionnel de catalogage partagé, relevant du back office : il n'a pas vocation à se substituer à un outil public de signalement et d'interrogation en ligne des données, ni à une bibliothèque numérique, même s'il a vocation à les alimenter via des connecteurs. Il ne permet pas non plus la gestion dynamique des collections physiques, bien qu'une nouvelle fois il importe que cet outil de production puisse irriguer des catalogues et outils locaux.

II. Besoins fonctionnels identifiés et modalités de mise en œuvre

Au cours de ses quatre premières réunions, le groupe de travail a inventorié un ensemble de besoins fonctionnels auxquels le futur outil doit répondre et les a regroupés en sept blocs principaux :

1. gestion des autorités,
2. gestion des contrôles,
3. gestion des droits,
4. convivialité, facilité d'installation et maintenance,
5. conversion des données,
6. gestion des unités physiques,
7. formation initiale/assistance/boîte à outils.

À la suite de la réunion du comité de pilotage du 19 janvier 2015, le groupe s'est ensuite interrogé sur les modalités de mise en œuvre de ces besoins fonctionnels. L'objectif du projet « outil national en EAD » est de mettre à disposition de l'ensemble des bibliothèques françaises dotées de fonds de manuscrits et d'archives un outil commun, **en le concevant dans sa globalité opérationnelle**. Le groupe de travail a considéré qu'il était nécessaire de prendre en compte l'existant et la grande hétérogénéité des utilisateurs potentiels, tant du point de vue des moyens techniques, financiers et humains, que des attentes et des besoins. Il a donc préféré procéder par étapes, à l'image d'un escalier que l'on gravit marche à marche, chacun à son rythme. Ont ainsi été définies trois étapes (ou marches) différentes :

1. la possibilité, pour toutes les bibliothèques, d'avoir au moins accès à un éditeur simple et pratique, quel qu'il soit, permettant de réaliser des notices descriptives dans le respect des normes de l'EAD et répondant aux critères énoncés par le groupe de travail. **Cette première étape ne prévoit pas à ce stade de mise en réseau d'un outil.**
2. **la mise en réseau des bibliothèques participantes** autour d'un système partagé d'identifiants et d'un référentiel commun d'autorités, avec **automatisation du transfert des données produites vers le CCFr** et convergence renforcée quant aux bonnes pratiques et aux évolutions techniques des outils.
3. **la mise en place d'un outil de production unique**, géré et maintenu par une structure **disposant de moyens et d'une équipe dédiée** au service des participants sur l'ensemble du territoire national.

Le groupe de travail tient à rappeler ici que si le niveau 1 de développement de l'outil peut répondre aux besoins des bibliothèques aujourd'hui dépourvues d'outil (pour l'essentiel des bibliothèques municipales mais également des petites structures de l'ESR pour lesquelles Calames apparaît surdimensionné et des entités associatives), il ne répond à lui seul en rien à la dynamique suscitée par le lancement de ce projet interministériel, et ne se conçoit qu'en lien avec une évolution parallèle et convergente des outils de production des deux grands réseaux existants (BnF et ABES), aujourd'hui plus proches du niveau 2.

Sans pour autant le considérer comme un objectif commun durablement suffisant, le niveau 2 peut en revanche apparaître comme une étape cohérente au regard de la lettre de mission du groupe de travail en dessinant les contours d'un outil national multipolaire par la fédération de bases s'inscrivant dans des environnements et contextes distincts, le renforcement de la

coordination des moyens destinés à asseoir l'EAD en bibliothèque, la confirmation du CCFr dans son rôle d'agrégation plus que de centralisation au sein d'une hypothétique base de données unique.

La partie qui s'ouvre ici vise donc à reprendre les différents besoins fonctionnels identifiés et la façon dont ces derniers peuvent se décliner en fonction du degré de développement et de centralisation de l'outil (3 degrés ou marches). On a par ailleurs noté pour mémoire le positionnement actuel des outils existants et leurs éventuelles adaptations nécessaires.

1. Gestion des autorités

En préambule, rappelons que le principe selon lequel l'outil national de production en EAD doit pouvoir s'appuyer sur un **référentiel national d'autorités** a été acté par le comité de pilotage.

- Au premier niveau de développement de l'outil, cette fonction devra permettre le contrôle des autorités créées à partir des référentiels existants, qu'ils soient nationaux ou internationaux mais également locaux, dans la mesure où des services d'interrogation adéquats sont disponibles.
- Le passage au second niveau dépend notamment de la mise en place d'un référentiel national d'autorités, ainsi que d'un protocole d'alimentation de cette base, qui pourrait être alimentée au départ par les données de la BnF et de l'ABES. Dans ce scénario, l'ensemble des utilisateurs de l'outil peuvent se rattacher à des notices déjà existantes ou proposer d'en créer de nouvelles.
- Le troisième niveau permet quant à lui soit d'exporter, soit de travailler directement dans la base de production nationale, avec des identifiants uniques.

Rappel de l'existant et évolutions à envisager : Depuis 2008, les outils de l'ABES et de la BnF fournissent des solutions facilitant la normalisation des points d'accès EAD (fonction de copier-coller des autorités BnF depuis l'outil de catalogage ; module d'interrogation et de liage automatique aux autorités Sudoc côté ABES, ayant cédé la place en 2011 à un interfaçage avec l'application IdRef, qui comme Calames et tout en étant enchâssé dans l'outil Calames Prod, permet la modification et la création de métadonnées en ligne et en réseau).

Reste toujours en suspens la question des référentiels locaux, alimentés et gérés par certaines bibliothèques (par exemple les grandes bibliothèques municipales comme la bibliothèque municipale de Lyon). Par ailleurs, un réservoir international agrégeant plusieurs référentiels est également disponible (VIAF) et doit être pris en compte dans notre réflexion, de même que les réflexions autour de l'EAC menées par les archivistes.

2. Gestion des contrôles

L'EAD est un format XML. Il importe donc de s'assurer que les instruments de recherche produits dans ce format sont à la fois bien formés (répondant aux règles de base du XML) et valides (conformes à la DTD), ce que propose en standard la totalité des éditeurs XML. Cette

fonction de contrôle par la norme et par le format a pour corollaire les contrôles liés à l'import/export de données (*cf. infra*).

En outre, les instruments de recherches produits doivent également correspondre à un standard de description commun, application des bonnes pratiques de l'EAD en bibliothèque (*cf. supra*).

- Au premier niveau de l'outil, avec une gestion très décentralisée, il semble difficile de contraindre les catalogueurs. Le groupe de travail propose donc de prévoir la possibilité de conserver en local des fichiers non conformes aux bonnes pratiques (mais bien formés et valides), tout en proposant aux catalogueurs des dispositifs d'alertes et/ou une bibliothèque de modèles descriptifs prévoyant un squelette d'inventaire ou de notice obligatoire et non modifiable. Ce contrôle de conformité doit pouvoir être fait sur les documents créés par le catalogueur mais également sur les documents importés.
- Au deuxième niveau de développement de l'outil, le groupe de travail propose la mise en place d'un sas d'analyse et de correction des données importées ainsi que d'un système de prévisualisation.
- Le dernier niveau, très centralisé, est celui qui assure le mieux cette gestion des contrôles de conformité, dans la mesure où il implique la mise à disposition d'une équipe dédiée à la gouvernance de l'outil et à un accompagnement scientifique coordonné.

Rappel de l'existant et évolutions à envisager : un important travail d'harmonisation et de préconisation a été mené depuis le début des années 2000 autour des bonnes pratiques de l'EAD en bibliothèques et fournit un cadre à la mise en œuvre de ces contrôles. Le groupe de travail propose donc la mise en place d'une instance de suivi (qu'il reste à définir) afin d'assurer l'évolution de ces bonnes pratiques, en prenant notamment en compte l'éventuelle mise en œuvre de l'EAD3 et à même d'instruire de futures modélisations de données à la jonction des travaux archivistiques (modèle ICA) et bibliographiques (FRBR).

À ce jour, seul Calames dispose d'un dispositif d'alerte en cours de saisie (non renseignement de certains attributs signalés par soulignement) et d'un outil de contrôle non contraignant (feuille de style, visiocontrôle), assurant une prévisualisation et une vérification de certaines bonnes pratiques.

3. Gestion des droits

- Le premier niveau de développement de l'outil n'implique pas vraiment une gestion des droits tant que l'hébergement reste local. Le groupe préconise néanmoins de prévoir au minimum 2 niveaux : administrateur/catalogueur.
- Le deuxième niveau de développement de l'outil (mise en réseau) ne concerne que les établissements très complexes, très hiérarchisés, comme la BnF. Dans ce cas précis, le groupe considère que l'organisation du travail interne à chaque établissement doit permettre de régler la question de la gestion des droits.
- Enfin, au troisième niveau, la gestion des droits doit également prévoir le verrouillage des fragments d'instance en cours de modification.

Rappel de l'existant et évolutions à envisager : Calames et PiXML disposent déjà de leur propre système de gestion des droits (à deux niveaux). L'outil Calames est actuellement le seul à permettre un accès authentifié en ligne extérieur à l'établissement d'hébergement.

Une mise en réseau de l'outil (niveau 2) nécessitera la mise en place d'un système de gouvernance assurant cette gestion des droits pour l'ensemble des partenaires du réseau, de même que dans le cas d'un outil centralisé (niveau 3).

4. Convivialité, facilité d'installation et maintenance de l'outil

- Dès les premier et deuxième niveaux, il convient d'être le moins dépendant possible d'un quelconque système d'exploitation (Windows, Linux, etc.). À ce stade de sa réflexion, le groupe de travail souhaite privilégier une installation la plus légère possible pour les établissements, de type application web, accessible à distance (i.e. si possible utilisable sans disposer des droits d'administration sur le poste informatique). La mise en place de masques de saisie et d'une bibliothèque de modèles descriptifs (*cf. supra*) devra contribuer à une meilleure prise en main de l'outil par les usagers peu familiers de l'XML/EAD, sans pour autant annuler la possibilité pour les utilisateurs le souhaitant (la majorité des producteurs de données actuels) d'un encodage « direct » susceptible de permettre un signalement plus détaillé et précis.
- Le niveau 3 quant à lui implique *de facto* l'existence d'une équipe dédiée et bien identifiée pour ces questions de maintenance. La forme de son intervention reste toutefois à définir : hot line ? Mailing list ? La mise en place d'une telle équipe peut d'ailleurs être envisagée sous une forme plus « légère » aux niveaux inférieurs.

Rappel de l'existant et évolutions à envisager : les outils Calames et PiXML partagent depuis l'origine le même éditeur XML (XMAX v5.0, plugin web de l'éditeur XMetal développé par la société JustSystems et distribué en France par la société Sollan) et souhaitent s'affranchir des contraintes liées à ce logiciel (obsolescence croissante de certaines technologies Microsoft). L'ABES et la BnF disposent d'équipes informatiques dédiées à même d'assurer les opérations de suivi et de maintenance.

Une solution susceptible d'intéresser les deux réseaux a été développée par la MSH de Caen dans le cadre de Biblissima autour du logiciel XML Mind-XML Editor (XXE). Les possibilités qu'offre XXE de multiplier les « vues » sur tout l'éventail allant du code XML au formulaire simplifié ont notamment porté le groupe à prendre en compte cet éditeur dans les propositions formulées en partie III.

5. Conversion des données en amont et en aval de l'outil

- Dès le premier niveau, il convient de prévoir l'import de notices déjà existantes dans les formats suivants : EAD 2002 et EAD3, mais aussi, autant que possible les formats MarcXML et Marc ISO 2709 (Unimarc, Interarc), les données tabulées (CSV), MODS, Dublin Core. Il convient de prévoir, en parallèle, l'export des données produites aux formats suivants : EAD 2002 et EAD3, Marc XML et Marc ISO2709, MODS, Dublin Core, données tabulées (CSV), HTML, PDF, METS, TEI, RDF.

Même si le groupe de travail a souhaité prendre en compte une typologie très large de formats, c'est surtout l'import/export des formats Marc vers l'EAD qui constitue une priorité au regard des gisements de données informatisés existant actuellement dans le réseau des bibliothèques du CCFr.

Enfin, dès ce niveau, une **fonction d'export des données vers le CCFr**, la plus simple possible, doit être impérativement prévue, selon des modalités à définir en concertation avec le département de la Coopération de la BnF.

- Le passage au deuxième niveau implique la mise en place, au niveau des outils de diffusion (Calames, BAM, CCFr) d'une méthode de chargement et de publication des données produites. Ce chargement devra bien évidemment permettre la prise en compte d'opérations de contrôle (*cf. supra* point 2) ; pour les BM, le CCFr aura des droits sur leurs instances.

Rappel de l'existant et évolutions à envisager : la majorité de ces conversions relèvent soit d'un traitement par macros (ex. des données tabulées), soit de feuilles de style employables au moyen de la plupart des éditeurs XML, qui les uns ou les autres nécessitent analyse et expertise.

Il convient donc de réfléchir à une solution globale de l'archivage pérenne des données produites, puisque l'ABES et la BnF assurent d'ores et déjà la sauvegarde de leurs propres données et que le CCFr, dont la mission prioritaire est de permettre une recherche fédérée, stocke néanmoins, *de facto*, 450 000 composants EAD qui font bien entendu l'objet d'une sauvegarde.

6. Gestion des unités physiques

Le groupe de travail considère que ce besoin fonctionnel peut faire l'objet d'un traitement spécifique, **hors outil national de production en EAD**. Il suggère toutefois que des spécifications techniques suffisamment claires soient fournies aux fournisseurs de SIGB, afin que ceux-ci puissent travailler à l'interopérabilité entre leurs produits et l'outil national de production en EAD.

Pour ce qui a trait à l'interopérabilité avec des bibliothèques numériques, le groupe de travail propose la mise en place de solutions d'import et/ou d'export associant *a minima* identifiants de niveaux descriptifs et liens <dao>.

Rappel de l'existant et évolutions à envisager : dans le cadre des évolutions à apporter à PiXML, la BnF travaille à l'intégration de cette fonctionnalité de gestion des unités physiques (unités de conservation), utiles notamment pour la gestion de chaînes de numérisation (l'implémentation automatisée des <dao> après traitement dans Gallica est déjà en place) et la préparation des transferts de la phase 2 du projet de rénovation du quadrilatère Richelieu.

7. Formation initiale/assistance/boîte à outils

Les besoins identifiés doivent également prendre en compte la formation des usagers au nouvel outil, quel que soit le niveau de développement de l'outil. Seules les modalités de la formation et de l'assistance évolueront en fonction des degrés de développement de l'outil.

- Au premier niveau, le groupe de travail propose la mise à disposition d'un didacticiel en ligne ainsi que la formation à l'outil national à l'ENSSIB (formation continue) ou dans les structures en régions (CRFCB). La question de la gratuité de cette formation, y compris pour les agents territoriaux, doit être posée. Pour le réseau des BM, l'offre de formation pourrait être conçue dans le cadre du réseau des pôles associés en région.
- Au deuxième niveau, le réseau des structures régionales pour le livre pourrait gérer la formation au plan administratif, avec une assistance scientifique et technique, à destination des BM. La BnF et le réseau Calames assureraient de leur côté la formation des usagers de leur réseau.
- Au troisième niveau, la formation pourrait être assurée par l'équipe dédiée à l'outil, en faisant appel à des renforts éventuels auprès des grands établissements relevant du MESR ou du MCC.

Pour ce qui touche aux questions d'assistance technique, le groupe de travail suggère de la confier à un prestataire extérieur aux niveaux 1 et 2. Le niveau 3, le plus centralisé, permet de disposer d'une équipe dédiée qui pourra assurer cette fonction.

Rappel de l'existant et évolutions à envisager : la formation à l'EAD et ses outils reste aujourd'hui multiple, la BnF, le CCFr et l'ABES assurant des formations liées à leurs outils ou contextes respectifs ; l'ENSSIB et certains CRFCB proposant de leur côté des formations (initiale et continue) au catalogue en EAD. La mise en place d'un outil national de production peut être l'occasion d'unifier ou du moins de mieux coordonner cette offre de formation.

Par ailleurs le Guide des bonnes pratiques actuellement disponible en ligne offre un modèle et une base à développer pour un éventuel outil didacticiel/forum à destination des catalogueurs. Le réseau Calames dispose également d'outils d'assistance propres (liste de diffusion, guichet d'assistance ABES, plateforme de formation à distance...).

Questions de gouvernance et d'organisation

Après cette description des besoins fonctionnels et de la façon dont ils peuvent être satisfaits aux différents niveaux de développement de l'outil, il convient de réfléchir aux questions de **gouvernance** et d'**organisation** propres à chacune des marches envisagées.

- Le premier niveau, s'il semble plus simple à mettre en œuvre, requiert toutefois des opérations de maintenance et de suivi qui peuvent être confiées soit au concepteur de l'outil, soit à un prestataire chargé simplement de sa maintenance, des actions de formation et de dépannage à distance dans le cas d'une application Web.

Il semble également important de mettre en place, dès ce niveau, un dispositif de suivi représentatif des différents partenaires impliqués dans la mise en œuvre d'un outil national, tant au niveau stratégique qu'au niveau technique et fonctionnel. Ce dispositif pourrait prendre la forme d'un petit réseau d'« experts » dont le modèle se baserait sur l'expérience de travail du réseau Calames.

- Le deuxième niveau, qui inclut une dimension de mise en réseau plus importante, notamment sur la question de la gestion des droits (*cf. supra* point 3), des liens aux référentiels d'autorités (*cf.* point 1) et de remontée des données au CCFr (*cf. supra* point 2), implique *de facto* la création d'une équipe de correspondants/techniciens, pas nécessairement localisés en un même lieu, mais capables de collaborer de façon renforcée au service du réseau constitué.
- Le troisième niveau, qui correspond à un outil centralisé, implique *de facto* la constitution d'une équipe dédiée, en charge de tous les aspects de gestion de l'outil national depuis son implémentation, la formation des catalogueurs, les opérations de maintenance et d'évolution de l'outil jusqu'au suivi de l'exposition des données.

Rappel de l'existant et évolutions à envisager : l'ABES et le CCFr disposent déjà d'une expérience en matière de gestion de réseau de catalogage. Pour autant les approches diffèrent en matière de structuration du réseau, de modèle économique (gratuité pour le CCFr, conventionnement avec facturation pour l'ABES), de modes de soutien à la production de données. Il convient donc de réfléchir au modèle organisationnel et économique souhaité par les tutelles pour le nouvel outil de production en EAD.

III. Scénarisations envisagées

1. Agencement des besoins

La richesse des besoins fonctionnels énoncés par le groupe de travail doit être regardée comme la traduction de l'attention particulière envers la diversité des ensembles documentaires et des formes d'indexation qui caractérisent, notamment, la pratique de l'EAD en bibliothèque. La bonne compréhension des interrelations entre besoins fonctionnels appelle également une nécessaire désambiguïsation du terme d'« outil XML/EAD ». Du cœur vers la périphérie, on peut identifier plusieurs couches aussi bien techniques qu'organisationnelles :

- un éditeur XML,
- un environnement de gestion d'un ensemble de fichiers XML et d'habilitations, enchâssant cet éditeur (du type PiXML ou Calames Prod) ou en lien étroit avec cet éditeur,
- des modes d'interaction (plus ou moins directs¹²) avec un outil d'indexation et d'interrogation des données,
- des modalités d'échanges et d'interopérabilité des données, qui dépendent notamment des combinaisons entre les trois couches précédentes,
- des questions de gestion de réseaux d'utilisateurs professionnels et d'institutions partenaires, et plus largement de gouvernance.

Les relations entre ces différentes couches diffèrent en fonction de l'horizon stratégique fixé au-delà de la première marche et induisent un certain nombre de tensions et de difficultés d'articulation au sein des besoins fonctionnels identifiés par le groupe, notamment :

- l'existence de besoins pouvant relever aussi bien d'un éditeur XML que de l'environnement de gestion voire d'une base de données (notamment : besoins n°1 et n°5), de même que l'existence de besoins ne pouvant relever que d'un environnement de gestion (notamment : besoins n°3 et n°6). Seules les catégories de besoins n°2 et n°4 peuvent être regardées comme pouvant relever strictement d'un « simple » éditeur XML. Ce dernier groupe de besoins pourrait certes être traité communément par tous les acteurs du GIT et indépendamment des autres besoins, mais sans que cela ne trace de voie en direction d'un outillage plus complet et plus partagé pour l'EAD en bibliothèque.
- le souhait pour les utilisateurs d'une installation la plus légère (répondant en cela au besoin n°4, et sachant que cela ne correspond qu'à peine au cas de Calames Prod, solution actuelle la plus évoluée à cet égard : un paramétrage par un administrateur de

¹² Calames-Prod, développé dans la logique d'une édition et d'une gestion par fragments, propose ainsi une intégration plus poussée que la BnF qui a recours à un espace tampon entre le plugin XMAX et l'environnement de gestion PiXML à proprement parler .

poste restant nécessaire lors du premier accès à l'outil) et la moins coûteuse possible (principe dans l'esprit duquel a été conçu l'outil Calames et le modèle économique de son réseau, et allant jusqu'à une exigence de gratuité de services dans le modèle de coopération BnF) d'une part, et le souhait d'un outil d'édition en XML et accessible en ligne d'autre part (or, aucun des éditeurs XML connus du groupe de travail, et notamment ceux proposant des plugins web, n'est à la fois gratuit et pourvu de fonctionnalités suffisamment riches)

- l'ambition réticulaire et nationale du futur outil est peu compatible avec un mode de production unique : le souhait de disposer de l'éventail le plus large possible, depuis le formulaire de saisie guidée jusqu'à la visualisation de l'arbre et du code XML, découle notamment de la variété des acteurs et producteurs actuels et futurs de l'EAD en bibliothèque. Par ailleurs, bien qu'une convergence croissante s'observe entre ces deux acteurs, la gestion de réseaux qui définit les activités de l'ABES se distingue en plusieurs points de l'approche adoptée par la BnF dans le cadre de sa politique de coopération (modèles économiques, modalités de soutien à la production ou à l'enrichissement de données, dispositifs de formations et d'assistance...) : or ces différences d'approches peuvent influencer significativement sur les directions à prendre au fil des trois niveaux de scénarisation.

2. Solutions écartées par le groupe de travail

Tout en définissant un cadre et des besoins communs, le groupe de travail, en amont du Comité de pilotage du 19 janvier 2015, s'est employé à examiner un panel (nécessairement non exhaustif) de logiciels actuellement utilisés dans différentes structures (*cf. supra* §3.1). Ces produits ont été exclus des scénarisations finales du fait qu'ils ne remplissent pas certains des critères définis comme essentiels par le groupe : notamment la gestion des contrôles XML (besoin n°2), le stockage XML natif et la possibilité d'éditer aussi bien en mode formulaire que du code XML (besoin n°4), la facilitation du liage à des référentiels d'autorités (besoin n°5). Très significativement, aucun des outils examinés n'a été conçu ni déployé dans la logique de réseaux d'institutions (besoin n°3). Par ailleurs, ni la vigilance à laquelle a invité le Copil au sujet de l'open source (qui soulève des questions de développements et de maintenance informatiques, voire de modèle économique, que le groupe de travail ne peut ni instruire ni trancher), ni le pragmatisme adopté par le groupe devant le manque de précisions relatives aux moyens communs, n'autorisent à privilégier des solutions propriétaires ou hypothéquant par trop l'autonomie des maîtres d'ouvrage vis-à-vis de l'évolutivité et de la maintenance du futur outil.

Une mention particulière doit être faite des produits suivants, bien que d'autres aient également été examinés (*cf. supra*) :

- ICA-ATOM¹³, solution libre et gratuite¹⁴ de catalogage EAD, EAC et EAF développée avec l'appui du Conseil international des archives, profilée sur les normes de

¹³ <https://www.ica-atom.org/>

¹⁴ Les développements sont cependant assurés par la société canadienne Artefactual Systems.

catalogage américaines (DACS), a retenu l'attention du GIT par sa capacité à gérer certains imports / exports (imports en XML et CSV, exports en Dublin Core) ainsi que la gestion d'unités physiques. En revanche, la production de données strictement encadrée par un formulaire ne traduisant qu'imparfaitement la richesse de l'XML/EAD semble peu compatible avec la richesse et la diversité des cas descriptifs rencontrés en bibliothèques et avec les pratiques actuelles de nombres d'établissements, sauf à se contenter d'un simple « signalement ».

- Mnesys Archives¹⁵, développé par la société française Naoned, comporte différents modules dont un « module de classement » composé d'une grille de saisie ISAD(G), qui ne correspond pas, en l'état, à l'éventail des possibilités de saisie des données EAD requis par le groupe de travail. La problématique de l'évolutivité d'un outil propriétaire dans la perspective d'un fonctionnement en réseau a aussi fortement conditionné le positionnement du groupe.
- La société Ever a fourni, sur demande du CCFr, un travail d'estimation et une démonstration de ce que pourrait être un éditeur XML permettant de faire de la correction et de la saisie contrôlée en EAD accolé à son progiciel, Flora. Cette proposition – chiffrée à environ 92 000 € –, forgée à partir d'un éditeur XML déjà développé dans le cadre d'une autre suite logicielle, semble répondre à certaines fonctionnalités de la première marche (besoins n°2 et 4 remplis ; éditeur full web ; éventuel liage à des référentiels externes). Elle n'a cependant pas été retenue par le groupe de travail dans la mesure où, totalement lié au progiciel et donc au CCFr, l'éditeur ne servirait donc qu'une partie de la communauté, et, à moins de développements spécifiques supplémentaires particulièrement coûteux, ne pourrait pas répondre à la totalité des besoins fonctionnels identifiés par le GIT.

3. Examen de l'éditeur XXE

Lors du lancement de notre réflexion, l'absence de dispositif de production EAD pour les établissements CCFr-CGM et Palme, ainsi que l'inadaptation croissante du plugin XMAX utilisé par les environnements de production BnF et ABES, entraient en résonance avec les réflexions et développements en cours au sein de l'Equipex Biblissima, autour de la production et de l'exploitation de (méta)données patrimoniales et notamment EAD.

Dans le souci de fournir aux chercheurs de son consortium un outil efficace, totalement orienté utilisateurs, permettant de créer des descriptions de manuscrits ou des éditions électroniques de sources, l'equipex Biblissima, se fondant sur l'expérience du pôle document numérique de la MRSH de Caen en la matière, a décidé de réaliser un éditeur XML centré sur deux grammaires XML : l'EAD et la TEI. Le développement *ex nihilo* d'un tel logiciel dépassant le cadre temporel et budgétaire du projet, Biblissima a fait le choix, comme l'avait fait auparavant la MRSH de Caen, de fonder ses développements sur un cœur logiciel non libre mais facilement accessible : XMLMind XML Editor. Une négociation avec la société

¹⁵ <http://www.mnesys.fr/logiciel-gestion-archives/>. A ne pas confondre avec le package « Bibliothèque numérique ».

Pixware, éditrice du logiciel, a permis d'une part d'acquérir une licence de type développeur permettant si nécessaire de modifier le logiciel lui-même et d'y adjoindre, dans les meilleures conditions techniques possibles, un certain nombre de modules complémentaires (développés par Biblissima en licence libre : raccordement à des bases distantes, édition collaborative de documents XML par exemple) et d'autre part de pouvoir diffuser gratuitement au sein des neuf établissements membres du consortium (dont la BnF) autant de licences du logiciel que souhaité, à partir du moment où celui-ci ne permet de créer ou d'éditer que de l'EAD ou de la TEI (pas d'autre grammaire XML).

Cet éditeur d'arbre XML présente le gros avantage d'être entièrement orienté auteur, ce qui veut dire qu'il ne permet à aucun moment de produire du code invalide (contrairement à beaucoup d'autres éditeurs qui, bien qu'ils soient équipés de dispositifs d'alerte, permettent cette opération). En outre, la création d'interfaces de saisie et de structuration spécialisées par type de modèle de données (ou par type de manuscrits par exemple) mais aussi par type d'opérations (analyse, annotation, etc.) est particulièrement aisé et permet donc d'adapter facilement le logiciel à différents besoins.

Le logiciel existe en plusieurs versions : client lourd et applet java. Si les développements réalisés dans le cadre de l'équipement d'excellence devraient fonctionner avec les deux versions du logiciel, client lourd et applet, seul le premier est utilisé par les membres du consortium.

4. Scénarios retenus par le groupe de travail

Le groupe de travail a finalement identifié deux grandes directions possibles en vue du développement en trois temps d'un outil national :

- d'une part, l'hypothèse de la création d'un outil *ex nihilo*, qui présente l'avantage de répondre à la totalité des besoins fonctionnels identifiés, tout en supposant de construire la convergence des deux outils et réseaux préexistants vers une nouvelle structure les englobant. Cette hypothèse « idéale » doit être considérée comme une aide à la décision pour choisir l'outil existant le plus susceptible de servir à la réalisation des différentes marches proposées.
- d'autre part, l'hypothèse de développements réalisés par la BnF ou par l'ABES à partir de leurs outils de production EAD actuellement en service, par réplique et enrichissement de PiXML ou Calames Prod. L'expérience des projets passés et les moyens actuels nécessaires à la maintenance et aux développements rendent ces scénarios plus aisément chiffrables en termes de coûts humains et financiers que l'hypothèse *ex nihilo*. Ces deux scénarios, présentés l'un par l'ABES à la réunion du GIT du 05/05/2015¹⁶, et l'autre par la BnF le 16/09/2015¹⁷, permettent tous deux de tirer profit d'acquis techniques et organisationnels, et de répondre prioritairement au besoin d'un outil de catalogage partagé exprimé par les établissements CCFr-CGM.

¹⁶ Cf. annexe 4.

¹⁷ Cf. annexe 5.

En revanche, aucun d'entre eux ne saurait être directement mis en œuvre, hors cadre stratégique approprié, sans prendre le risque d'une démarche projet asymétrique et préjudiciable au cadre interministériel congénital à l'EAD en bibliothèque. Les termes de différenciation entre ces deux scénarios résident essentiellement dans :

- l'existence d'un outil déjà « orienté réseau » (Calames Prod). En ce sens, les scénarisations Calames, tout en cantonnant l'intervention de l'ABES en « back office », reviennent à investir dans son expérience de gestion de réseaux¹⁸ et à consolider son rôle dans la fourniture de produits et de services mutualisés pour la production en réseau et en ligne de données documentaires.
- la mission nationale de coopération qu'assure la BnF et dont le CCFr est l'une des traductions. En ce sens, la scénarisation PiXML s'insère dans la perspective d'un élargissement d'usage des outils BnF hors de ses murs et auprès de ses réseaux de coopération. L'expérience de la BnF en matière de mise à disposition de son outil d'accès et de moissonnage distant des archives de l'internet pourrait être mise à profit pour un certain nombre de questions techniques et organisationnelles.

Ces trois scénarios présentent tous des points faibles et des points forts que le GIT a tâché de lister. Tous posent des questions techniques, financières et stratégiques que le groupe ne pouvait et n'avait pas vocation à explorer dans le cadre qui lui était imparti, et qui n'en sont pas moins des clés indispensables à la formalisation de pans entiers de scénarisations. Ainsi les estimations de coûts humains et financiers pour un outil créé de toutes pièces sont-elles très largement dépendantes d'options dont le groupe de travail ne peut présenter ici qu'un éventail.

¹⁸ Voir l'étude Calames 2014 (disponible en ligne : <http://www.abes.fr/Media/Fichiers/Calames/Calames-bilan-enjeux-et-perspectives-2006-2014>).

Degré d'avancement	Scénarios	Points forts	Points faibles
Marche 1	Calames	<p>La plupart des fonctionnalités déjà existantes dans l'outil de catalogage Calames peuvent être de nouveau mises à profit :</p> <ul style="list-style-type: none"> -imports de documents EAD possibles, nombreux exports XSLT en EAD/Marc/HTML (mais pas les expositions OAI-DC et RDFa qui tiennent à l'interface publique de Calames) : cf. besoins #1 et #6 -outil orienté réseau, gestion des habilitations et verrouillages en édition sur une pluralité de fichiers XML (2 niveaux administrateurs / établissements) : cf. besoin #3 -restriction de l'encodage à la seule DTD EAD et aides à la saisie et au contrôle selon les bonnes pratiques françaises : cf. besoin #4 -liage automatique aux autorités du référentiel IdRef (également amendable en ligne) déjà possible : cf. besoin #5 -réemploi de web services existants et adaptables, pour le moissonnage des données Calames par le CCFr -recul de huit ans et expérience de l'ABES en matière de production de données en réseau dans le cadre de Calames. 	<ul style="list-style-type: none"> -Dans l'option d'une répllication la plus directe possible et engageant le moins de développements possible (formulaires de saisie) : dépendance au seul OS Windows (navigateur Internet Explorer) liée à XMAX, avec des risques d'obsolescence liés aux évolutions Microsoft. Des dysfonctionnements dans l'usage d'XMAX sont d'ores et déjà identifiés à partir de Windows 8, et difficilement solubles. -La solution de moissonnage mise en place pour les données Calames reste non standard, <i>ad hoc</i> -L'installation des droits d'accès à l'outil nécessite d'être administrateur du poste.
	PiXML	<ul style="list-style-type: none"> -Appui sur l'expérience de la BnF dans l'utilisation de son outil, il bénéficiera d'un réseau d'utilisateurs interne à celle-ci. -Maintenu par le DSI de la BnF, il bénéficiera également des développements planifiés et en cours de déploiement par celui-ci (changement du plugin XMAX, nécessitant des droits d'installation sur poste). -Plus grande fréquence à venir des moissonnages BnF par le CCFr. -Le liage automatique aux autorités BnF est d'ores et déjà planifié pour PiXML BnF, la version répliquée pourrait donc aussi en bénéficier. L'éventuel raccordement à un/des autres référentiels est envisageable. 	<ul style="list-style-type: none"> - PiXML est aujourd'hui un outil strictement BnF, très intégré à son SI (gestion des envois en numérisation, et bientôt, dans l'interface publique BAM, gestion de la communication des UC), dont les développements ont été faits en interne, ce qui est à la fois un point fort et un point faible. Les paramètres internes à la BnF devraient être exclus de la version mise à disposition du réseau CCFr. - Outil actuellement tributaire d'XMAX (coût licences, installation nécessaire du plugin sur poste, instabilités encore signalées qui pourraient s'aggraver en cas de connexion distante). - Les modules d'import et export ne comptent actuellement que l'EAD.
	Ex nihilo	<p>La création d'un outil ex nihilo permet le paramétrage à façons et notamment des développements pour le raccordement aux bases d'autorités existantes dès la marche 1, de masques de saisie adaptés aux besoins des catalogueurs, d'un outil conçu pour être indépendant des systèmes d'exploitation, de feuilles de transformation intégrées permettant les conversions dès la marche 1, etc.</p> <p>Dans l'hypothèse d'un recours à XXE, Biblissima serait en position de "passeur" potentiel d'une solution (répondant aux besoins 2 et 4) à adapter à au moins trois cas de figure (BaseX à l'origine, PostGRE pour Pixml, SQL Server pour Calames et "l'outil CCFr" le cas échéant).</p>	<p>Il est très difficile de paramétrer un éditeur permettant de répondre à la totalité des besoins listés dès la marche 1 <i>et</i> qui soit compatible avec l'ensemble des bases de données utilisatrices. L'interaction de l'outil avec les catalogues existants est également une inconnue de taille.</p>
Marche 2	Calames	<ul style="list-style-type: none"> -Accès optimisé pour le CCFr aux données EAD de son réseau, grâce à de nouveaux web services. -Marche impliquant l'émergence d'un référentiel national d'autorités opérationnel pour la récupération et la création de notices d'autorités : or IdRef, déjà interfacé avec l'outil Calames, en est la préfiguration la plus avancée. -En cas d'adoption de l'applet XXE : capitalisation possible des expériences MRSH Caen (formulaires de saisie) et ABES (gestion de réseaux) 	<ul style="list-style-type: none"> -En cas de pluralité durable de bases de données (marche 2 comme marche terminale), on n'aboutit pas à un outil national complet. Une fédération de bases de données n'est cependant pas nécessairement un point faible : dans l'optique d'un EAD plus orienté vers les données et plus interopérable, ce peut même être un avantage. - Non-généricité des développements, à l'égard de l'association éditeur XXE + base de données SQL Server -Des limites dans les possibilités d'exports (via SQL Server)

	PiXML	<ul style="list-style-type: none"> - Gestion des contrôles : il est déjà prévu de faire évoluer le système de visualisation de PiXML. - Gestion des droits : PiXML distingue déjà entre administrateur et catalogueur. 	<ul style="list-style-type: none"> - La question de la création / modification d'autorités (actuellement via une application de catalogage interne) reste entière. - Gestion des contrôles : actuellement les seuls contrôles se font au niveau de XMAX, non de PiXML. - Convivialité : PiXML est actuellement trop dépendant d'un système d'exploitation ; les masques sont insuffisamment ergonomiques et très délicats à gérer. - Gestion des droits : actuellement PiXML autorise tout agent à intervenir sur les IR produits par d'autres services que le sien, ce type de contrôle est donc à développer dans le cadre d'une mise à disposition au niveau d'un réseau.
	Ex nihilo	-Echafauder un système d'identification pérenne des données convergeant vers le CCFr	-Les environnements de catalogage qui enchâssent l'éditeur d'usage général ne peuvent sans doute pas être identiques dans tous les contextes (ex. définition des arborescences Calames par les catalogueurs, suivi des workflows Gallica à la BnF...)
Marche 3	Calames	<ul style="list-style-type: none"> -Entrepôt commun de données EAD, pour des traitements communs à l'ensemble des inventaires EAD en bibliothèques et pour un travail concerté sur la modélisation, l'interopérabilité et la valorisation de ces données - Aboutir à une interface d'interrogation professionnelle des données EAD, et non uniquement à une interface de consultation/modification document EAD par document EAD 	- Des co-développements ou l'intégration commune de développements sont une <i>terra incognita</i> vis-à-vis des collaborations entre BnF et ABES.
	PiXML	<ul style="list-style-type: none"> - Gestion des autorités : PiXML gère déjà l'attribution d'identifiants uniques aux IR et aux composants ; ces identifiants uniques seront en outre bientôt pérennes - Gestion des droits : actuellement PiXML gère déjà le verrouillage des composants en cours de modification 	- Les spécificités organisationnelles (tailles des collections) et informatiques de la BnF risquent de contraindre l'outil.
	Ex nihilo	<p>Une centralisation aboutie de l'outil de production en EAD suppose une maintenance réactive, et une adaptabilité forte aux contextes des différents utilisateurs (workflows différenciés de publication).</p> <p>L'éventualité d'un outil partagé et centralisé de production en EAD en bibliothèque pose la question de la convergence des pratiques. La difficulté serait d'harmoniser les contrôles de manière à ce qu'ils soient suffisants, en marche 3, pour satisfaire tous mais pas trop contraignants pour servir tous les utilisateurs.</p> <p>-Un outil ex nihilo paramétrable de manière à ce qu'un EAD davantage encore harmonisé et respectueux des bonnes pratiques soit produit à cette étape.</p>	Un outil centralisé ne peut que très difficilement prendre en compte les besoins de type gestion des unités de conservation, propres à chaque établissement.

5. Coûts comparés

Scénarisations Calames

En marche 1 *stricto sensu*, ce scénario peut être concevable à moyens constants, à savoir les équipes Calames (env. 3 ETP) et CCFr-Mss (env. 1,5 ETP) actuelles, et considérant que la première décision apportée est d'ordre organisationnel (rapprocher les équipes dédiées aux réseaux de l'EAD en bibliothèque). L'outil Calames serait alors ouvert au réseau CCFr-CGM moyennant les adaptations suivantes :

- possibilité de créer des masques de saisie dans le plugin XMAX (via XFT) : env. 60-80 j/H, soit environ **18.000 €**. Ce développement est cependant peu compatible avec l'option d'une refonte de l'outil Calames autour d'un nouveau plugin (développée en marche 2) et présente des risques quant à la pérennité du cœur d'édition XML
- interrogation des autorités BnF depuis l'outil Calames (via les services VIAF) : 20 j/H, soit environ **5.000 €**
- nouveaux web services CCFr –CGM : 30 j/H, soit environ **7.000 €**

Cependant, le scénario Calames ne prend tout son sens qu'en atteignant plus directement la marche 2. L'outil Calames Prod est refondu – ou, dans la perspective du dispositif répliqué à l'usage du CCFr, l'outil est directement conçu – autour d'un nouveau plugin web d'éditeur XML (hypothèse d'une applet java XXE notamment) : 200 j/H, soit environ **55.000 €**. A cela doivent être ajoutés :

- des besoins en **ressources humaines** parallèles à ces (re-)développements, et liés à la nécessité de maintenir techniquement l'outil Calames « historique » durant cette période
- des **coûts de licence développeur** à prévoir, en fonction de l'éditeur XML choisi (pour information : 15.000€ dans le cas d'XXE)

Scénario PiXML

Scénario de niveau marche 2 dès le départ. Estimé à 135 jours de développements pour un coût estimé à **91.000 €** environ. Il s'agit d'estimations basses. Ce coût est valable pour les marches #1 et #2 puisque PiXML permet déjà de se raccrocher aux autorités BnF et en 2016, de le faire sans passer par le module de catalogage interne BnF.

Gestion des autorités : marche impliquant la mise en place d'un référentiel national d'autorités
Convivialité : développement d'un système de gestion d'une bibliothèque de masques de saisie, améliorations graphiques et ergonomiques prises en compte dans le chiffrage.

Scénario ex nihilo

Le coût financier et humain n'a pu être estimé par le groupe de travail, qui n'a pas les compétences pour ce faire. À titre informatif, la société Ever, prestataire du progiciel Flora utilisé par le CCFr, a chiffré à **92 000 €** (estimation basse) le « simple » raccordement d'un éditeur XML existant adapté EAD avec masques de saisie et prévisualisation (une sorte de « demi-marche 1 ») au CCFr, réponse directe et ciblée à l'une des priorités à l'origine de la constitution du GIT qui ne peut néanmoins répondre qu'à une partie des attendus fonctionnels validés par le comité de pilotage du 19/01/2015.

Plus généralement, au vu des coûts de développement appliqués aux outils existants, le coût de développement d'un outil entièrement créé pour répondre aux besoins fonctionnels identifiés par le groupe de travail ne pourra qu'être nettement plus élevé que les scénarios Calames ou PiXML. Un co-développement BnF-ABES pour un outil et une base EAD interministérielle pourrait être une solution ambitieuse, mais au regard de laquelle nous ne disposons d'aucun précédent. Quant aux économies d'échelle en termes de maintenance, elles ne se sauraient se dessiner qu'à l'horizon d'une marche 3 réunissant un outil, une base, une équipe et un dispositif réseau totalement unifiés (perspective qu'il n'appartient pas au groupe de privilégier davantage qu'une fédération renforcée de ressources demeurant indépendantes).

6. Conséquences stratégiques et questions à résoudre

Quel que soit le scénario finalement retenu, il est nécessaire d'envisager l'instauration, dès la marche 1, d'un **mode de gouvernance interministériel**, garant notamment de la poursuite et de l'approfondissement de la coopération entre les différents acteurs. En marche 2, la problématique est notamment liée aux formes que prendra le futur référentiel national d'autorités. En marche 3, se pose la question stratégique de la mise en place de la structure responsable de l'outil et celle liées à l'articulation avec la publication des données (donc le CCFr). Enfin, il convient de s'interroger communément sur le statut de l'EAD en bibliothèque, à la charnière entre données bibliographiques et archivistiques.

Scénario Calames

En marche 1, mise en place d'une répartition possible des rôles entre ABES et CCFr :

- Équipe ABES en charge des conventionnements avec les établissements relevant du MCC, ainsi que de la maintenance et de l'évolution de l'infrastructure technique de « l'outil CCFr ».
- Équipe CCFr en charge de l'accompagnement à la production de données (contrôles qualité, soutiens aux rétroconversions...) et aux formations des établissements.

L'établissement BnF, qui souhaite conserver PiXML à moyen terme, n'est a priori pas directement concerné par le dispositif de réutilisation et d'adaptation de l'infrastructure Calames.

En marche 2, mise en place d'un cofinancement ESR et Culture (licences éditeurs, infrastructures serveurs supplémentaires) pour les deux outils Calames et CCFr et d'une gouvernance commune BnF-ABES d'une base nationale d'autorités.

Dans l'hypothèse de l'usage de l'applet XXE, la MRSH-Caen et Biblissima seraient en position de facilitateurs potentiels (contours d'une collaboration à définir).

En marche 3, déploiement à envisager au sein de la BnF d'un dispositif éprouvé par les réseaux CCFr et Calames. La possibilité d'une migration de l'outil national hors de l'escarcelle informatique de l'ABES (livraison à la BnF/CCFr) doit rester envisagée (après phase de capitalisation côté ABES, basculement en marche 3 côté Culture ?).

Scénario PiXML

Dès la marche 1, la BnF devient maître d'ouvrage pour l'outil partagé. La mise en place d'une gouvernance interministérielle permettant de prioriser et d'arbitrer les évolutions de l'outil, dans une double logique de besoins réseaux et établissements, est d'autant plus nécessaire. Le

rôle de la BnF comme prestataire pour le CCFr en marche 1 et 2 est à instruire, l'implication future de l'ABES dans cet outil et le bénéfice que pourrait retirer Calames des développements BnF pour son application partagée à déterminer par le comité de pilotage. La BnF souhaite-t-elle se donner les moyens de fédérer autour de son outil un réseau large Culture et ESR ? Le catalogage en EAD est-il suffisamment prioritaire pour que des moyens soient alloués par les deux tutelles à long terme pour poursuivre non seulement la maintenance mais également les développements de l'outil ?

Scénario ex nihilo

En marche 1, la question se pose du recours à un prestataire avec pour condition des développements open source, ou du choix d'une équipe dédiée ? Quel partage de compétences, quel profit tirer des expériences existantes ?

En marche 2, se pose la question de la genericité des développements de base sur le plugin XML, et de l'éventail des coûts d'adaptation permettant l'accès à diverses bases de données. Si l'appropriation par les différents réseaux est faite dès la marche 1, cela sous-entend-il des adaptations techniques à chacun de ces réseaux et donc une divergence rapide entre ceux-ci ?

Conclusion

La réflexion conduite à l'occasion de cette étude de faisabilité nous conduit à dresser les constats et à poser les questions suivantes, qui doivent être pris en compte quel que soit le scénario retenu au final pour l'outil de production en EAD.

- Par-delà l'hétérogénéité des situations des établissements concernés par le futur outil, les feuilles de route propres à chacun portent à accorder des degrés de priorité variables aux différents besoins énoncés, à définir des orientations différentes aux actions en cours ou à venir sur les différentes couches d'outil.
- Dans ce contexte, il est intéressant de se demander jusqu'où doit aller l'exigence d'open source exprimée par certains partenaires.
- La prégnance de 2 voire 3 environnements de production et de gestion de fichiers en EAD semble empêcher de raisonner en termes d'outillage commun, à moins
 1. d'envisager sérieusement des co-développements BnF-ABES
 2. d'envisager sérieusement un rattachement du réseau CALAMES à la BnF ou vice versa.
- L'objectif à se fixer est-il une base de données unique ? Le CCFr sera-t-il le gestionnaire de cette base ? Avec quels moyens ? La via media d'une centralisation fédérative (marche 2) est-elle à exclure ?
- La dédramatisation appelée autour de la question de la production en XML/EAD (besoins 2 et 4) semble également nécessaire sur la question des bases de données, la manière de stocker, de traiter, d'exploiter les données en EAD
- Il importe donc de ne pas séparer les problématiques outils et données.
- Il est essentiel d'avancer sur la question de l'outil de production aux côtés de la communauté des archivistes et de faire en sorte de converger avec eux sur la question des référentiels, de la normalisation et de la modélisation.
- Enfin, il est essentiel de poursuivre et d'accentuer la collaboration entre les équipes déjà dédiées à l'EAD, tant à la BnF et à l'ABES que dans les établissements (BM, BU, etc.), autour des questions de structuration, de gestion et d'approche réseaux, de contrôle qualité de l'EAD en bibliothèques, etc. Un suivi régulier, par ses différents acteurs, des étapes de la mise en oeuvre d'un outil de production partagé est nécessaire et ne pourra que renforcer cette collaboration.

Annexes

Annexe 1 – Déclaration d'intention du 17/06/2014 « Concertation nationale autour d'un outil commun de catalogage en EAD »

Depuis 2002, les catalogues décrivant les collections de manuscrits et d'archives de la Bibliothèque nationale de France et des bibliothèques publiques et universitaires ont été en grande partie informatisés. Cette réalisation s'est effectuée en particulier grâce à la numérisation et la structuration au format EAD (Description archivistique encodée) de grands ensembles catalographiques : *Catalogue général des manuscrits des bibliothèques publiques de France*, catalogues du département des manuscrits de la BnF, *Répertoire des manuscrits littéraires français du XXe siècle*.

Ces chantiers d'informatisation se sont accompagnés de travaux de normalisation : recommandations DeMarch sur la description des manuscrits et archives modernes et contemporains¹⁹ et Guide des bonnes pratiques de l'EAD en bibliothèque²⁰. Les données issues de l'informatisation sont consultables à travers trois catalogues principaux : Calames²¹, BnF archives et manuscrits²² et le Catalogue collectif de France²³ (CCFr). L'intégralité de ces données est par ailleurs interrogeable simultanément à travers le CCFr.

Pour les besoins de ses départements spécialisés (Manuscrits, Arts du spectacle, Audiovisuel, etc.) la BnF a développé en 2006-2007 l'outil de catalogage PiXML, qui arrive bientôt à obsolescence. De son côté, l'Agence bibliographique de l'Enseignement supérieur fournit aux établissements du réseau Enseignement supérieur le logiciel de catalogage Calames, lui aussi créé vers 2006-2007²⁴. Les bibliothèques publiques en revanche ne disposent pas, via le CCFr, d'une interface de catalogage. Par ailleurs, l'absence sur le marché de logiciels dédiés au catalogage au format EAD représente l'obstacle majeur au signalement des collections de manuscrits des bibliothèques publiques et à l'enrichissement du *Catalogue général des manuscrits*.

En-dehors du monde des bibliothèques, mais en lien avec un projet de signalement d'une collection de manuscrits et d'ouvrages imprimés (projet *Bibliothèque virtuelle du Mont-Saint-Michel*), la Maison des Sciences de l'Homme de Caen a développé depuis 2010 un prototype de logiciel ergonomique dédié au catalogage en EAD et à l'édition de textes en TEI.

¹⁹

<http://www.bivi.fonctions-documentaires.afnor.org/content/download/19537/135613/version/4/file/demarch.pdf>

²⁰<http://www.bonnespratiques-ead.net/>

²¹<http://www.calames.abes.fr/>

²²<http://archivesetmanuscrits.bnf.fr/>

²³<http://ccfr.bnf.fr/portailccfr/jsp/index.jsp>

²⁴L'ABES vient de mener un audit sur ce logiciel afin d'en apprécier les forces et faiblesses : <http://www.abes.fr/Media/Fichiers/Calames/Calames-bilan-enjeux-et-perspectives-2006-2014>

Ces travaux de développement sont maintenant repris par l'équipe d'informaticiens de l'équipex Biblissima, auquel appartient la MSH de Caen.

Désireuse de fournir une solution à ses catalogueurs, consciente également des contraintes des bibliothèques publiques, et mise au fait du développement en cours d'une solution par Biblissima du fait de son appartenance à l'équipex, la BnF a souhaité réfléchir à une solution globale. Dans son nouveau contrat de performance, elle s'est engagée à développer un outil de production en EAD et à s'attacher à « la mutualisation avec les autres institutions concernées, dans l'optique d'un futur outil national de production en EAD qui constituerait la brique logicielle indispensable à l'enrichissement et au développement du catalogue national des manuscrits. » Un tel outil pourrait être partagé par les bibliothèques publiques et leur permettrait de s'approprier le format tout en évitant au CCFr d'avoir à retraiter systématiquement les données qui lui sont adressées.

L'élargissement aux bibliothèques universitaires a paru représenter une évolution intéressante et la mutualisation d'un outil EAD a donc été évoquée lors de réunions entre le ministère de la Culture et le ministère de l'Enseignement supérieur, en particulier dans le cadre du renouvellement de la convention entre l'ABES et la BnF.

A l'issue d'une première réunion le 17 juin dernier au ministère de la Culture et de la Communication, ces différents partenaires se sont mis d'accord sur une démarche commune en vue de réaliser une étude de faisabilité d'un outil national, et ont approuvé la déclaration d'intention ci-dessous.

Déclaration d'intention

Le ministère de la Culture et de la Communication (DGMIC/Service du livre et de la lecture),

le ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche (Département de l'information scientifique et technique et réseau documentaire),

la Bibliothèque nationale de France,

l'Agence bibliographique de l'Enseignement supérieur,

et l'équipement d'excellence Biblissima

s'accordent sur :

le constat que la description en EAD des manuscrits, archives et fonds structurés est nécessaire pour une bonne connaissance des collections patrimoniales françaises et pour fournir aux usagers des bibliothèques les informations nécessaires à leur travail

le constat parallèle que l'offre logicielle actuelle n'est pas satisfaisante et pose un obstacle au développement souhaitable du signalement des collections

l'intérêt de mener une étude sur la faisabilité d'un outil commun de catalogage en EAD, propre à être utilisé par toutes les bibliothèques françaises qui le souhaiteraient

la création d'un groupe de travail composé de représentants des partenaires précités et de leurs réseaux, auquel est assignée la mission suivante :

Faire un **état des lieux des solutions techniques** actuelles pour le catalogage en EAD

Définir un **noyau consensuel de besoins fonctionnels** attendus d'un outil commun de catalogage en EAD : le groupe se concentrera sur les questions logicielles pour la description en EAD (et accessoirement en EAC), en prenant en compte l'évolution probable des schémas et la nécessité de pouvoir gérer simultanément les formats EAD2002 et EAD3. Le groupe réfléchira également à l'opportunité d'inclure des fonctionnalités de gestion des unités physiques et de communication des documents.

Élaborer des scénarios et options détaillant les principales spécifications du futur outil.

Réfléchir aux **dimensions économiques du projet** : ressources humaines, budget de développement, répartition des charges, etc.

Produire un **rapport** à l'issue de l'étude de faisabilité.

Le calendrier de travail prévoit une publication du rapport de cette étude vers la fin du premier trimestre 2015.

Annexe 2 – Composition du groupe de travail « Outil national EAD »

1. Composition du groupe de travail :

Ministère de la Culture et de la Communication / Direction générale des médias et des industries culturelles / Service du livre et de la lecture/ département des bibliothèques :
Bureau du Patrimoine : Pierre-Jean RIAMOND

Ministère de l'Enseignement supérieur et de la recherche/ DGESIP-DGRI/ Département de l'information scientifique et technique et réseau documentaire :
Julie LADANT

Bibliothèque nationale de France :
Direction des collections/ Département des Arts du spectacle : Patrick LEBOEUF
Direction des services et des réseaux/ Département de l'information bibliographique et numérique : Vincent BOULET
Direction des services et des réseaux/ Département de la coopération : Camille POIRET

Agence bibliographique de l'Enseignement supérieur :
Département des services aux réseaux : Jean-Marie FEURTET

Equipex Biblissima :
Maison des sciences de l'Homme de Caen : Pierre-Yves BUARD
BnF/Direction des collections/Département des manuscrits : Matthieu BONICEL

Représentants de la lecture publique :
Médiathèque de Poitiers: Florent PALLUAULT
Bibliothèque municipale de Lyon : Jérôme SIRDEY

Représentants du réseau Calames :
Bibliothèque Mazarine : Patrick LATOUR
ADBU : Frédérique JOANNIC-SETA

2. Personnalités auditées par le groupe de travail :

- Ministère de la Culture/ Direction générale des Patrimoines/Service interministériel des archives de France/Sous-direction de la politique archivistique :
 1. Claire SIBILLE
 2. Alice MOTTE
- Archives nationales : Direction de l'appui scientifique : Pascal TANESIE

Annexe 3 – Liste des réunions

N°	Date de réunion	Ordre du jour
1	29 septembre 2014	Définition de la méthode de travail – rappel des attendus et du calendrier
2	4 novembre 2014	Définition des besoins fonctionnels attendus
3	3 décembre 2014	Définition des besoins fonctionnels attendus
4	9 janvier 2015	Définition des besoins fonctionnels attendus – Audition des représentants du SIAF et des AN
5	10 février 2015	Bilan du comité de pilotage du 19 janvier – Définition de la typologie attendue des utilisateurs
6	11 mars 2015	Définition des marches de déploiement de l’outil national
7	30 mars 2015	Examen croisé des marches de déploiement et de la typologie des usagers
8	5 mai 2015	Travail sur les scénarios
9	22 juin 2015	Travail sur les scénarios
10	Juillet 2015	Tests techniques
11	16 septembre 2015	Elaboration du rapport final

Annexe 4 – Sept années d’encodage dans Calames Evolutions d’un réseau de production en EAD (2008-2014)

Note pour le GT Outil national EAD (réunion du 11/03/15)

Chiffrer les activités et l’effort de production de métadonnées

Chiffres généraux sur les profils de données EAD présentes dans la base (février 2015) : se référer au tableau statistique général du réseau pour plus de précision

Base Calames originelle (fin 2007) : env. 140.000 <c> issus des rétroconversions CGM et Palme.

A l’automne 2014, franchissement de la barre du demi-million de <c> publiés dans Calames. La base Calames est extrêmement parisienne (88% des documents décrites sont conservés à Paris ou petite couronne), et les **cinq plus gros établissements de Calames** (Institut de France, MNHN, BLJ Doucet, Académie de Médecine, BDIC) représentent à eux seuls, non seulement les 2/3 de la base publique, mais aussi les 2/3 de l’effort général de production de données et d’encodage fourni durant sept années. Le phénomène est peut-être paradoxal (le besoin de réseau aurait pu naître d’une dispersion d’êts.), mais est en réalité un facteur clé dans la naissance de l’initiative Calames (cohésion d’un groupe d’êts. moteurs).

Si l’on ramène les efforts de chaque établissement déployé (49 êts. déployés fin 2014) au temps écoulé depuis lequel il a accès à l’outil Calames, on obtient une moyenne générale **d’environ 2000 <c> créés par établissement et par an** (depuis le MNHN, producteur de plus de 20.000 <c> par an en moyenne, jusqu’à une dizaine d’êts. déployés et néanmoins absolument inactifs). « L’effet réseau » escompté pour Calames n’est pas allé jusqu’à atténuer le grand écart originel entre des besoins de travail collaboratif sur un long terme, et des besoins de saisie ponctuels et minimes en EAD.

Les ¾ des 1030 fichiers EAD publiés dans Calames sont dans la responsabilité d’¼ des établissements Calames (i.e. les 12 plus importants RCR indexés, sur 66 RCR indexés au total).

[On dénombre: 838 fichiers EAD comportant un contenu descriptif en <dsc>, 104 fichiers liés, 92 « fichiers-feuilles » (où seuls des hauts niveaux descriptifs sont publiés), 100 fichiers maîtres ou sous-maîtres.]

Comment sont créées les données EAD ?

Du fait de l’architecture intégrée de Calames et d’une volonté initiale d’attacher strictement l’intégration au réseau Calames avec un recours à l’outil d’encodage, la quasi totalité du crû de données publiées (+ 360.000 <c> créés dans Calames et publiés en sept ans) est passée par l’outil Calames, avec toutefois des nuances entre :

- créations complètes d’IR en EAD via Calames, généralement en transposant des IR existants
- reprise d’IR Word (avec des problèmes de copier-coller et de mises en forme des données quelquefois), générés en fin de chaîne de traitement archivistique, préalablement au signalement
- catalogage « de but en blanc » : plus fréquent avec le temps, pour traiter des collections de documents, voire des fonds d’archives de tailles modestes ou moyennes,

et considérant que l'encodage de données en EAD se confond désormais davantage de l'acte de créer un instrument de recherche.

- import de données EAD préexistantes (créées dans oXygen, parfois Arkheia ou ICA-Atom), suivi d'un contrôle dans Calames pour un affichage et une indexation corrects dans Calames, le cas échéant liées à IdRef et/ou davantage précisées dans leur encodage (c'est du moins l'exigence posée par l'ABES).

Dans une minorité de cas, la présence des données dans Calames doit s'articuler avec l'existence d'une autre interface de publication locale (Pleiade notamment).

Il est envisagé d'élargir le conventionnement Calames au cas d'établissements ne recourant pas à l'outil de catalogage (pas de facturation XMetal). Ce qui laisse ouvertes les questions du contrôle des bonnes pratiques et du liage aux référentiels.

Calames est un outil intégré, cependant l'acte de publier y est dissocié de l'acte d'encoder (enregistrer une modification dans le code EAD n'entraîne pas automatiquement la publication de cette modification). Fin février 2015, on compte dans la base Calames près de 100.000 <c> présents mais jamais publiés (soit l'équivalent d'un quart des 380.000 niveaux descriptifs à la fois créés via l'outil Calames et publiés dans la base). Cette notion de différentiel concerne aussi les mises à jour de données : une proportion non négligeable (5 à 10%) des données publiées au moins une première fois sont en réalité sur l'établi, à un instant T, et en attente d'une nouvelle publication. Au total, on peut estimer qu'un quart des fichiers et données EAD sont « en chantier », et que cette proportion s'est avérée être assez constante depuis plusieurs années.

A la lumière d'une nouvelle procédure statistique mise en place à l'automne 2014, voici une analyse détaillée de trois mois d'activité de catalogage dans Calames (novembre 2014 - janvier 2015) :

- 246 fichiers EAD modifiés au moins une fois dans Calames, soit un peu plus de 15 % (un sixième) du total (1584) des fichiers présents.
- Sur ces 246 fichiers, 113 n'ont fait l'objet de modifications qu'au cours d'une seule journée ; 75 fichiers de 2 à 5 « reprises quotidiennes » ; 35 fichiers de 6 à 10 reprises quotidiennes ; et 23 fichiers à plus de 11 reprises quotidiennes (jusqu'à 39 et même 45 reprises pour les fichiers (fonds Doucet Bernard Noël et Mazarine Renier). Chiffres à mettre en regard d'environ 60 jours d'activités possibles en moyenne sur trois mois.
- Près de la moitié des actions de créations / modifications sur <c> ont été le fait de trois RCR seulement (BIF, BLJD, INHA). La plupart des 28 RCR ayant agi ont modifié 1 ou quelques fichiers EAD ; quelques-uns ont cependant eu une action sur leur arborescence également (52 fichiers modifiés, souvent ponctuellement, par le MNHN). Environ 15% des fichiers modifiés (37 sur 246) ont fait l'objet de plus de 1000 « modifications-c » au cours des trois mois²⁵ (jusqu'à 16000 pour un fonds BLJD).

Principes et caractéristiques d'un réseau de catalogage

En-dehors de l'existence d'un outil intégré qui est en lui-même facteur d'échanges entre établissements (à travers les formations initiales ou relais, ainsi que des impacts variables sur les chaînes de traitements documentaires), les dénominateurs communs à chaque

²⁵ C'est une des difficultés de ce décompte : si un établissement ne modifie qu'un seul niveau descriptif, mais ayant de nombreux descendants, ces derniers seront eux aussi considérés comme modifiés.

établissement Calames tiennent aux bonnes pratiques EAD, aux calendriers administratifs des conventionnements ABES-établissement (procédures de déploiements et de renouvellements de conventions triennales, facturation annuelle de licences XMetal, liste de diffusion), ainsi qu'au dispositif de soutien financier aux conversions rétrospective. Le réseau Calames présente une forte particularité par rapport aux réseaux Sudoc ou Thèses de l'ABES : ni le catalogage ni le catalogue ne peuvent être partagés, puisque les descriptions sont uniques, intimement liés à des contextes de production et de conservation, et portent sur des documents ou ensembles uniques ; cependant Calames s'est doté d'un fonctionnement collégial spécifique (réunion périodique du GT Calames, héritier de la structure projet).

En février 2015, la liste de diffusion Calames compte environ 115 inscrits en établissements (dont 50 correspondants Calames : référents automatiquement inscrits en cas de déploiement). Ces 115 personnes ne correspondent pas forcément aux producteurs de données eux-mêmes (ex. : un correspondant Calames non-catalogueur sera abonné à la liste de diffusion ABES, mais pas tel contractuel embauché plusieurs mois).

Le soutien de l'ABES aux conversions rétrospectives occupe une place majeure parmi les facteurs expliquant la part importante d'évolutivité des données de la base. De 2009 à 2014, un montant de 629.179 euros a été distribué à 21 établissements Calames. Selon les années, les activités de catalogage dans l'outil Calames ont été soutenues à hauteur de 30 à 50% par ces sommes. Actuellement, env. 100.000 € sont distribués chaque année et se traduisent presque exclusivement par des recrutements de contractuels ou vacataires, dont la stabilisation ou fidélisation est souvent problématique (d'autant que ce sont généralement des établissements engagés sur des chantiers pluriannuels qui se trouvent en position d'être durablement demandeurs).

La « part silencieuse » d'un réseau

Depuis l'origine de l'outil Calames en 2008, l'inactivité en production / modification de données EAD de la part d'établissements pourtant déployés est un phénomène structurel. En 2012, une enquête avait été menée auprès de neuf établissements déployés mais peu ou pas actifs. Depuis cette époque, certains se sont mis à encoder des données : il s'agissait donc pour eux d'une question de calendrier (anticipation du déploiement dans Calames par rapport à la priorisation ou à la capacité à mettre en branle les projets). Le coût de l'appartenance au réseau en cas d'inactivité (au minimum 200 ou 300 € / an pour une licence XMetal) n'apparaît que rarement comme problématique. Les principales explications mentionnées par les correspondants Calames interrogés sont le manque de temps et de moyens humains (dans 2/3 des cas), mais aussi des problèmes techniques avec l'interface de catalogage (dus à des proxys, au passage à de nouveaux environnements Windows, à la bonne gestion de certains caractères non latins), ou encore la concurrence avec d'autres projets (description de fonds imprimés et rétroconversions dans le Sudoc). A part une minorité d'établissements pour lesquels les documents rares et/ou anciens représentent une part très significative de leurs activités et projets, le patrimoine est par ailleurs loin d'être en bonne position dans les priorités des services documentaires. Ont aussi été exprimés certains besoins en formation (mieux maîtriser l'EAD) et/ou d'échanges accrus avec d'autres catalogueurs Calames : l'isolement (d'une personne référente dans un établissement, ou d'un établissement dans le réseau) est facteur de fragilité, d'ajournement ou d'abandon dans les projets.

Quelles marges d'accroissements ?

En 2011, le périmètre institutionnel théorique de Calames (pour peu qu'on considère aussi bien l'Enseignement Supérieur que la Recherche) était estimé à un peu moins d'une centaine d'établissements (ce chiffre pouvant gonfler en intégrant des services d'archives, ou en considérant que les collections photographiques et fonds hybrides aient lieu d'être traités en EAD). Le chiffre exact dépendra notablement de ce qu'on qualifie d'« établissement ». La définition la plus simple et formelle pose l'équation « un établissement Calames = une convention Calames »; or il arrive que plusieurs RCR et encore davantage d'équipes productrices de données soit couverts par une même convention : ex. Institut-Thiers, BULAC-EFEO, BIUS, centres Université Paris 1...)

Principales problématiques d'accroissements de Calames :

- du point de vue du nombre et de la diversité des partenaires : peu de SCD restent à « conquérir » et les principaux établissements CGM ou Palme (qui avaient là un point d'accroche pour s'intéresser à « leurs » données et intégrer le réseau) sont déjà dans le réseau. Les principales marges d'accroissement concernent les services d'archives et les organismes de recherche / laboratoires. Le prosélytisme parfois engagé en-dehors des périmètres CGM/Palme a eu des effets variables : de l'indifférence (FMSH) jusqu'à l'effet tache d'huile (ex. MMSH et adhésion progressive à Calames des éts. membres du Consortium HumaNum « Archives des Ethnologues »), en passant par un déploiement sans mise en branle de chantiers (Collège de France).
- en termes de périmètres documentaires : bien qu'il reste du travail à effectuer sur les collections de manuscrits initialement concernées par le CGM, ce sont les fonds d'archives privées et de la recherche (+ éventuellement, des archives institutionnelles) qui constituent les plus importants gisements. Ce sont souvent des fonds hybrides (comportant des documents non-manuscrits).

Quelques références supplémentaires

Voir le blog Calames, pour les statistiques 2014 :

<https://calames.wordpress.com/2015/01/19/les-statistiques-de-calames-en-2014/>

ainsi que pour une synthèse de la journée du réseau Calames organisée en mai 2014 :

<https://calames.wordpress.com/2014/06/06/retour-sur-la-journee-du-reseau-calames/>

Dans l'Etude Calames 2014, voir notamment p. 7-10 (« Calames, outil collectif et collaboratif » et « Faire évoluer le réseau ») :

<http://www.abes.fr/Media/Fichiers/Calames/Calames-bilan-enjeux-et-perspectives-2006-2014>

Annexe 5 – Scénariser un projet d’outil national EAD : le point de vue de l’équipe Calames

Les deux scénarisations qui suivent émanent d’une concertation au sein de l’équipe Calames : ils n’engagent que l’ABES, bien que leur élaboration ait notamment bénéficié des échanges avec les représentants d’établissements ESR siégeant en GT Outil national EAD.

« Calames Prod » désigne l’environnement de catalogage de Calames, accessible uniquement via le navigateur Internet Explorer : <http://www.calames.abes.fr/prod>

Scénario #1 : " Calames Prod adapté et mis à disposition hors ESR"

Ce scénario a connu une première évocation lors d’une réunion de travail entre ABES et DIBN le 5 février 2015. Cette solution serait la plus simple du point de vue de l’ESR, car conservant les fonctionnalités majeures de Calames et évitant l’éclatement de l’existant. La plus grande partie des développements Calames s’en trouvent réemployés et valorisés dans un périmètre général plus cohérent (celui de l’EAD en bibliothèques).

Sur le modèle de la duplication de base de données qu’offre déjà aujourd’hui la base de Formation, il s’agit de répliquer l’environnement de production (mais *pas* l’interface de publication) de Calames.

L’infrastructure reste globalement identique : base de données SQL Server²⁶, production en ligne via un plugin web d’édition XML, activité dynamique et en ligne sur les données (verrouillage des données en cours de modifications, gestion des données). Il semble de ce fait préférable d’éviter les dépenses et efforts qu’impliquerait un transfert de codes, de savoirs, d’expérience, et de serveurs vers la BnF (ou vers tout autre organisme que l’ABES).

La question de la visibilité des données Calames dans le CCFr a été tranchée en faveur d’un renvoi vers l’interface publique Calames : de la même manière, l’ABES ne se charge aucunement de publier les données, son rôle s’arrêtant à la mise à disposition directe des données auprès du CCFr (et/ou des établissements concernés). L’EAD est rendu moissonnable par le CCFr par web services après une forme simplifiée d’indexation dans SQL Server, ou par OAI-PMH.

Le modèle de conventionnement ABES est étendu aux établissements ressortant du MCC et souhaitant recourir à cet environnement de production adapté ; en échange, l’ABES pourrait bénéficier de ressources (à définir) à l’échelon interministériel pour consolider l’outil Calames.

Ce scénario permet de conserver (voire de favoriser) la possibilité d’un traitement de besoins exprimés à l’occasion du bilan Calames réalisé en 2014 : meilleure gestion des unités physiques et des informations à vocation interne, ou interface d’interrogation professionnelle et générale des données EAD type WinIBW (ce que l’outil de catalogage Calames n’est presque en rien, jusqu’à présent).

Scénario #2 : "Calames Prod réarticulé"

Il s’agit ici de construire un nouvel outil d’édition XML-EAD partagé à l’échelle interministérielle, et utilisable dans tous les contextes qui ont été recensés par le GT. Le

²⁶ SQL Server est certes une technologie Microsoft, mais en matière d’indexation et de gestion de données, nous ne voyons pas que les produits Oracle ou PostGre soient plus intéressants. (Par ailleurs, depuis quelque temps, Microsoft a initié une politique open source pour certaines briques Windows.)

recours à ce nouvel outil pourrait ainsi être rendu compatible avec une intégration au réseau Calames.

Pour l'ABES, cela revient à concevoir un nouveau moyen d'accès à la base de données Calames, moins intégré²⁷. Cela ne remet pas en cause, du moins dans un premier temps, l'existence de la solution intégrée déjà en place (Calames « historique »), mais vient en complémentarité.

Entre le nouvel outil EAD et les bases de données qui lui préexistent, il est possible que deux types d'interrelations soient nécessaires : un système rudimentaire *d'imports-exports* (cas du CCFr en principe), ou bien des *connecteurs* avec des bases de données stockant les fichiers EAD et permettant leur accès en ligne. Dans l'hypothèse d'un interfaçage « trop » simple (i.e. imports-exports) avec le CCFr, ce scénario #2 a plus de probabilité de voir le jour si la BnF souhaite elle-même adopter le nouvel outil pour son propre usage.

Le partage des fonctions serait le suivant :

- le nouvel outil EAD permet d'encoder selon la DTD et les bonnes pratiques, de faire usage d'un environnement ergonomique et de grilles de saisies adaptées à différents contextes, de contrôler et valider les données, d'importer et d'exporter/transformer les données, de lier à IdRef comme aux autorités BnF
- les données EAD conservées dans la base de données SQL Server de Calames ne pourraient être déchargées sur poste et ouvertes dans l'outil qu'après un verrouillage à l'échelle du fichier tout entier (et non plus pour un simple fragment, comme le permet Calames « historique »)
- un « sas » spécifique à cette nouvelle solution d'accès aux données Calames reste nécessaire pour la gestion des fichiers : organisation des arborescences Calames, détection des doublons, demandes de publications (complètes), accès aux traces de jobs. Ce « sas » serait proche de l'environnement Calames Prod actuel mais *sans* aucune possibilité d'édition /visualisation des données
- à la jonction de deux manières de produire de l'EAD, la base de données et l'interface publique Calames restent uniques.

Dans ce scénario #2, l'usage d'un outil EAD sur poste, et non accessible en ligne, pose encore des questions quant à la facilité d'installation (en fonction des règles DSI locales), et ouvre la voie à d'indéniables reculs fonctionnels : perte des fonctionnalités de travail partagé (qu'on a cependant reconnu comme sous-utilisées à l'occasion du bilan Calames), perte de la publication partielle (qui présente dans tous les cas des limites), abandon (probable) de pistes de développement d'une interface d'interrogation des données professionnelles pour Calames Production.

Quant à la question de l'outillage d'édition XML à privilégier, le parti pris du pragmatisme marque ces deux scénarios. Dans l'un ou l'autre des scénarios, il n'est pas évident qu'il faille exclure le recours à XMetal (choix historique et jusqu'alors commun à la BnF et à l'ABES), dont l'installation suppose certes d'avoir des postes sous Windows. Mais l'installation d'XXE suppose quant à elle l'installation d'une machine Java sur poste. XMetal permet comme XXE la création de masques de saisie et une customisation graphique reposant sur un emploi assez simple de fichiers css et xml. Les arguments possibles en faveur d'XXE seraient 1° de bénéficier de l'expérience et des développements MRSH Caen et Biblissima, 2° d'avoir une possibilité d'accéder aux sources de l'outil (négociations Biblissima), 3° d'acheter français (mais tout en se privant des communautés d'utilisateurs étendues dont XMetal ou oXygen

²⁷ En écho à la problématique des « déploiements sans recours à l'outil de production » (cas de figure initialement exclu dans le modèle du réseau Calames).

bénéficient). Dans le scénario #1, un recours à XXE dans « l'outil Culture » offrirait une opportunité de consolider l'environnement de production Calames, en collaboration avec la MRSH Caen qui a l'expérience d'XXE (customisation, ergonomie, liens aux autorités essentiellement).

Enfin, un outil véritablement partagé a aussi vocation à devenir unique, ou à ouvrir la voie vers un outil unique : mais à quelle échéance ? Ou au contraire, la coexistence entre les outils de production « historiques » (Pixml et Calames) et le nouvel outil national peut-elle être durable ?

Annexe 6 – Solutions présentées au groupe de travail par la BnF-CCFr le 16/09/2015



Direction des services et des réseaux
Département de la coopération

Quel outil pour la mise à jour du *Catalogue général des manuscrits* ?

Présentations au groupe de travail *Outil national EAD* des solutions examinées pour le CCFr

16 septembre 2015

Historique

La rétroconversion du Catalogue général des manuscrits (2003-2008) a concerné non seulement les bibliothèques municipales et universitaires, mais également la BnF et toutes les autres institutions françaises détentrices de collections de manuscrits (archives départementales, bibliothèques spécialisées, etc.).

Le Catalogue national des manuscrits, accessible exclusivement via le Catalogue collectif de France depuis 2013, se décompose en trois ensembles :

- BnF, avec BnF Archives et Manuscrits ;
- ESR, avec Calames ;
- Toutes les autres institutions, avec le *Catalogue général des manuscrits* du CCFr.

Le CCFr est donc **l'outil national de diffusion, proposant un accès fédéré et l'interrogation de la totalité des catalogues de manuscrits sus-cités.**

Or **le réseau du CCFr (qui représente 470 institutions) est le seul à ne pas disposer aujourd'hui d'une interface de production et correction de ces données**, bien qu'offrant comme indiqué une interface de diffusion via le CCFr.

En ce qui concerne son propre catalogue, la BnF (DIBN, DSI, DCO) travaille à terminer la refonte totale de l'interface de consultation du catalogue public BAM (pour début 2016) et améliore l'outil de production (PIXML), notamment par la gestion d'identifiants pérennes

(2015) et le remplacement de l'éditeur XML (2016/2017). Le DSI a assigné un budget au financement d'une étude comparative des éditeurs XML sur le marché dont les premiers éléments seront connus prochainement et qui sera mise à disposition de l'ensemble des acteurs représentés dans le groupe de travail.

Enjeux

Le département de la coopération de la BnF qui gère le Catalogue collectif de France, est animateur d'un **réseau de partenaires très large** par le biais des conventions passées, mais également par les différents services offerts à ce réseau. Le CCFr à lui seul représente un réseau conséquent (plus de 220 bibliothèques membres du réseau PIB, environ 180 bibliothèques présentes dans la base Patrimoine (dont 48 BMC), 470 institutions dans le CGM...), animé par de nombreux interlocuteurs en région.

Pour la BnF, dans le cadre de sa politique nationale de coopération, le premier enjeu est de **mettre à disposition du réseau du CCFr un outil permettant la mise à jour** (corrections des notices existantes et production de nouveaux inventaires) **des données qui ont été produites dans le cadre de la rétroconversion du CGM**. Cet enjeu est d'autant plus crucial que de nombreux partenaires existants demandent aujourd'hui à pouvoir procéder à cette mise à jour.

Le deuxième enjeu est de **permettre la récupération et la réutilisation des données**. Ce futur outil doit donc intégrer les normes existantes, notamment celle de produire impérativement un EAD conforme aux bonnes pratiques telles qu'elles ont été définies.

Le dernier enjeu pour le CCFr est de **récupérer facilement et dans un contexte contrôlé les données du CGM qui auraient été produites ou modifiées par les partenaires**.

Calendrier

Le CCFr s'est donné comme **période de mise à disposition d'un outil EAD pour les institutions présentes dans le CGM le terme du Contrat de performance de la BnF, soit la fin de l'année 2016**. Dans celui-ci, est évoquée la nécessité de "doter les personnels d'un outil de travail répondant aux fonctionnalités requises [pour la production en EAD] (...) qui constituerait la brique logicielle indispensable à l'enrichissement et au développement du catalogue national des manuscrits coordonné par le CCFr".

L'échéance de la fin de l'année 2016 permettra également de répondre à une **demande croissante des bibliothèques présentes dans le CGM de mettre à jour leurs données**,

données ayant été mises en ligne il y a plus de sept ans et corrigées seulement de façon ponctuelle depuis. Cette demande est fortement exprimée (notamment) par les partenaires des pôles associés régionaux.

Par ailleurs, le **développement d'un outil de production en EAD pour le CGM est inscrit au plan de développement logiciel 2016** de la BnF.

Méthodologie de travail

L'objectif premier du CCFr est de **mettre à jour systématiquement les instruments de recherche existants** en intégrant notamment les données disponibles (corrections et révisions); son deuxième objectif est la **production de nouvelles données** (nouveaux inventaires...) dans un cadre contrôlé et normalisé afin de mettre à disposition du plus grand nombre un catalogue complet et actualisé des collections de manuscrits et d'archives.

Jusqu'à maintenant, les corrections ont été marginales faute d'outil. En effet, bien que des formations soient dispensées par le CCFr, leur mise en application reste difficile et aléatoire ; de surcroît, l'intégration dans le CGM des données produites demande des ajustements nombreux, précis et chronophages.

Une méthodologie de travail à l'échelle régionale est en cours d'élaboration avec la région Champagne-Ardenne comme région-pilote. L'enquête qui y a été menée a fait ressortir le **très grand désir des collègues sur place de procéder à cette mise à jour, mais également leur besoin d'un outil adapté pour ce faire.**

La multiplication des projets régionaux de mise à jour du CGM nécessite d'une part une coordination ; d'autre part, une planification à moyen terme et enfin, une maîtrise des coûts. Il est nécessaire d'accompagner ces programmes par la mise à disposition d'un outil partagé. L'objectif final est, une fois ces opérations systématiques menées à bien, que les membres du réseau du CCFr se soient appropriés cet outil et poursuivent les mises à jour par eux-mêmes dans un cadre contrôlé au niveau national.

Propositions

En **marche 1** du scénario présenté par le groupe de travail, il s'agirait plutôt de proposer un outil qui ne soit pas un simple éditeur XML, mais bien **un programme de catalogage en EAD permettant une remontée automatisée** (autant que possible) **de données structurées, normalisées et interopérables dans le CCFr.**

Plusieurs solutions sont envisageables pour cette première étape :

- une solution développée par l'ABES, prenant appui sur Calames ;
- une solution PiXML, développée en interne par le DSI de la BnF, visant à ouvrir un accès distant aux partenaires du CCFr à ce programme de catalogage ;
- une solution développée par la société Ever, titulaire du marché BnF pour l'évolution et la maintenance du CCFr ;
- voire une autre solution, comme la création de toute pièce d'un programme de catalogage EAD à partir de briques logicielles existantes ou non.

Par ailleurs, la marche 3²⁸, qui correspond à un outil de catalogage partagé au niveau national, suppose une gouvernance centralisée à propos de laquelle aucun des acteurs aujourd'hui impliqués dans le projet ne s'est prononcé jusqu'aujourd'hui.

La rétroconversion du CGM n'a été que la première étape vers un catalogue national des manuscrits, aujourd'hui concrétisé par l'interface de consultation du CCFr. Des investissements raisonnables, allant dans le sens d'un outil national de catalogage en EAD pour une mise à jour du *Catalogue général des manuscrits* du CCFr, permettront de prolonger l'investissement initial en visant la production de données harmonisées et interopérables.

Ce dossier reste un point d'attention et de dialogue entre BnF, DCP, Culture et ESR qui ne doit pas obérer la nécessité de continuer à valoriser globalement l'utilisation de l'EAD dans les bibliothèques et à soutenir toutes les initiatives de collaboration en matière de normalisation et de bonnes pratiques de l'EAD.

²⁸ La marche 2 correspond à une étape intermédiaire au cours de laquelle l'outil (relativement) simple de la marche 1 serait (entre autres) connecté au futur référentiel national d'autorités.