

CALAMES

BILAN, ENJEUX ET PERSPECTIVES

Document mis en ligne le 17 juillet 2014

Sommaire

Objet et cadre de l'étude	2
Bilan de l'existant	2
Calames, outil de production : forces et faiblesses.....	2
Calames, une interface publique.....	5
Calames, outil collectif et collaboratif.....	7
Enjeux et perspectives.....	9
Faire évoluer le réseau	9
Remédier aux limites de l'outil induites par les choix technologiques initiaux	10
Assurer la cohérence, la mise à niveau et la qualité des données.....	10
S'adapter à l'évolution du contexte normatif et réglementaire	11
Développer l'interopérabilité.....	11
Favoriser l'exploitation des données	12
Rapprocher signalement et numérisation.....	13
Conclusion	15
Annexes statistiques.....	16
1. Statistiques par établissements : données publiées, créées, indexées (2007-2014).....	16
2. Evolution générale du nombre de <c> publiés dans Calames (2007-2014).....	18
3. Statistiques relatives aux points d'accès dans Calames	19
4. Statistiques de consultation (2008-2012)	20
5. Statistiques Calames Plus Images (chiffres mars 2014)	22

Objet et cadre de l'étude

Décidée lors d'une réunion du groupe de travail Calames (14/02/2014), la réalisation de cette étude a pour objet de dresser, six ans après son lancement, un bilan de Calames dans sa triple fonction d'outil d'encodage, d'outil de publication, et de réseau de production de descriptions de manuscrits et de fonds d'archives. Prenant en compte les évolutions intervenues depuis sa création ou se profilant dans un avenir proche, dans le domaine technique, dans le paysage normatif de la description encodée de manuscrits, dans la valorisation numérique des collections d'archives et de manuscrits, l'étude se veut également un examen des enjeux auxquels Calames est confronté. Quelles sont les perspectives d'évolution et d'adaptation des outils et du réseau Calames à un contexte national différent de celui qui a présidé à sa création ?

Bilan de l'existant

Calames, conçu en parallèle de l'opération interministérielle de conversion rétrospective en EAD du Catalogue général des manuscrits des bibliothèques publiques de France (CGM), a dès l'origine¹ été voulu non seulement comme un outil de consultation des catalogues encodés de manuscrits d'établissements relevant de l'Enseignement supérieur, mais aussi – et plus encore – comme un outil permettant la correction et l'enrichissement de ces instruments de recherche par ces mêmes établissements, appelés malgré le caractère unique des collections décrites à travailler en réseau.

De là découle le choix d'un outil collectif *intégré*, i.e. de production *et* de consultation d'instances EAD, différent en cela des choix effectués pour les catalogues encodés des bibliothèques d'établissements relevant de la Culture, tant pour ceux concernés par l'opération de conversion rétrospective du CGM (un outil de consultation – le Catalogue collectif de France (CCFr) – sans outil de production) que pour la Bibliothèque nationale de France² (un outil de production – PiXML – distinct d'un outil de consultation Pleade).

De manière originale par rapport à la plupart des catalogues, tant de manuscrits que d'imprimés, utilisés dans les institutions patrimoniales, Calames a été conçu, développé et est actuellement maintenu par les équipes informatiques de l'ABES en étroite collaboration avec des responsables de fonds patrimoniaux soucieux de recourir à l'outil le plus adapté possible aux usages constatés ou anticipés des usagers et à leur pratique quotidienne de signalement. Ce mode de fonctionnement particulièrement synergique a permis de faire évoluer un outil initialement conçu dans le cadre spécifique de la conversion rétrospective du CGM pour prendre en compte toute la diversité des collections patrimoniales (archives scientifiques et institutionnelles, documents imprimés, iconographiques ou audiovisuels liés à des fonds d'archives, collections organisées d'images ou d'objets, etc.).

Calames, outil de production : forces et faiblesses

L'interface de production de Calames recourt à un plugin web (XMAX) de l'éditeur XML XMetal, autour duquel ont été développées par l'ABES des fonctionnalités supplémentaires (volet arborescent d'accès aux

¹ Après commande de la sous-direction des bibliothèques et de l'information scientifique (SDBIS, Ministère de l'Enseignement Supérieur et de la Recherche) auprès de l'ABES début 2006, le calendrier du projet Calames fut le suivant : conception (mars-déc. 2006), développements et tests (janv.-nov. 2007), ouverture de l'interface en ligne (déc. 2007), mise en service de l'outil de catalogage (avril 2008).

² La BnF n'était partie prenante à la conversion rétrospective du CGM qu'au titre de l'Arsenal.

fichiers EAD et customisation de certaines fonctions XMetal)³. Le groupe projet Calames s'était prononcé en faveur d'XMAX du fait : 1° d'une concurrence quasi-inexistante des autres éditeurs XML en matière de plugin web⁴, 2° d'une pratique éprouvée au sein d'organismes patrimoniaux (dont les Archives nationales) ayant déjà réfléchi aux conditions d'amélioration de l'ergonomie, en vue d'une adaptation à des utilisateurs parfois peu familiers avec la manipulation de fichiers XML. En comparaison d'un éditeur XML classique, les principaux apports et spécificités de l'outil de catalogage sont de plusieurs ordres :

- Un accès par authentification permettant à chaque établissement de visualiser (lecture seule) l'ensemble des instances EAD présentes dans la base, et d'accéder en mode édition à ses propres données. Un fichier EAD peut être rendu accessible en édition à différents logins⁵.
- Une solution d'accès sur le web, indépendante du poste utilisé (dès lors qu'il a été correctement configuré) et du lieu de travail. La souplesse de ce dispositif permet à la plupart des établissements de ne déclarer qu'une ou deux licences XMetal auprès de l'ABES.
- Une sécurisation de l'édition en ligne par verrouillage des fragments (nœud courant et ses descendants). L'intérêt de ce dispositif est non seulement d'empêcher tout risque d'interventions parallèles et concurrentes, mais aussi d'offrir une possibilité de « versioning » à l'intérieur même d'une instance EAD⁶.
- Un volet polyvalent de gestion des instances EAD : réagencement de fragments, recherche dans les composants publiés⁷, fonctions d'imports et d'exports XSLT, de contrôle des doublons et de (dé)publication des instances EAD commandant directement la base de données à l'usage de l'interface publique. Une possibilité de « publication partielle » a été prévue pour éviter de republier des instances tout entières lorsque des modifications ponctuelles sont apportées à des fragments et parties de longues fraties⁸.
- Une possibilité d'interconnexion des instances EAD⁹ d'un même établissement grâce à des métadonnées *ad hoc* (ou « propriétés ») permettant de déclarer des chaînes de *liaisons* de fichiers, notamment lorsqu'un inventaire excède le poids maximal raisonnable¹⁰, et/ou des *inclusions* de

³ Parmi les principaux développements ayant permis d'améliorer l'ergonomie de saisie : caractère modulable des trois fenêtres de l'interface de production, convivialité et lisibilité du code EAD, renseignement par défaut de certains attributs et accès pour d'autres à des listes fermées de valeurs possibles.

⁴ Des essais de développement avaient également été entrepris autour du plugin Authentic de l'éditeur XML Spy (Altova), finalement non validés par le groupe projet.

⁵ Les identifiants d'accès à l'outil de catalogage ne sont pas attribués individuellement : ils correspondent à des groupes d'utilisateurs au sein d'un même établissement (charge aux correspondants Calames d'en répartir la divulgation au sein de leur établissement).

⁶ Cette fonctionnalité est dans la pratique très peu utilisée à l'échelle d'un même fichier XML : une répartition du travail dans le temps et entre les catalogueurs concernés, et/ou une subdivision des données à produire entre plusieurs fichiers temporaires voués à fusionner sur intervention du correspondant Calames, semblent être des solutions au moins aussi pertinentes.

⁷ L'extension de la fonction recherche aux données cataloguées mais non publiées a souvent été demandée à l'ABES ; il est d'autant plus difficile d'accéder à cette demande que le respect de l'intégrité native des fichiers EAD (*et non* un découpage par unités descriptives de type base de données relationnelles) a primé dans la construction des espaces de stockage alloués à l'outil de catalogage.

⁸ La fonctionnalité de publication partielle dispose de garde-fous (impossible de l'exécuter si le nombre de composants varie par rapport à l'existant dans la base, non plus que si le poids du fragment est tel que sa re-publication est aussi coûteuse qu'une re-publication intégrale) et présente de l'intérêt essentiellement pour des fraties longues au sein de fichiers peu arborescents (ex. : un <c> au sein de collections de manuscrits ou de photographies).

⁹ En complément de la présence, en hauts niveaux de fichiers, d'identifiants d'organismes responsables (RCR Calames) identiques, condition nécessaire mais pas suffisante pour définir la structuration d'une arborescence.

¹⁰ Pour toutes les bases XML se pose le problème de la gestion des fichiers pondéreux. C'est d'ailleurs en raison de sa faible capacité de charge que le recours à eXist-db (base nativement en XML) a été exclu au moment du projet Calames, au profit d'une base SQL Server. Le poids maximal recommandé des fichiers EAD a été fixé dans Calames à 3 Mo environ (à comparer au seuil d'alerte à l'édition de fragment fixé à 50 Ko, seuil fixé par prudence en 2008 mais qui aurait d'ailleurs lieu d'être rehaussé à 100 ou 200 Ko). En plus de fournir une solution à la publication

fichiers, dispositif qui permet d'étendre la notion d'arborescence à l'échelle de plusieurs fichiers EAD (un ou plusieurs fichiers « maîtres » commandant l'accès à une pluralité d'instruments de recherche pour refléter au mieux l'organisation structurelle des fonds d'un établissement).

- Une fonction de renseignement et de normalisation des points d'accès par un liage aux autorités de l'ABES, d'abord assuré par un module « Autorités Sudoc », puis par l'application IdRef à partir de 2011 (*cf. infra*). Les fonctions professionnelles de l'application IdRef (modifications et créations d'autorités), du fait du liage possible aux fichiers EAD, peuvent être considérées comme une extension particulière de l'environnement de production Calames.
- L'ABES, en lien avec le Groupe de travail Calames (*cf. infra*) et suivant les recommandations du Groupe national des bonnes pratiques EAD en bibliothèque, a mis en ligne un Manuel de catalogage à partir de 2009 et développé des modules de diagnostic et de contrôle de conformité à ce Manuel et aux exigences techniques de l'outil¹¹.

Cependant, un certain nombre de bugs (généralement non bloquants, et dont beaucoup sont également observés dans l'outil PIXML à la BnF) persistent. Depuis 2008, du fait notamment de la complexité et/ou de la fragilité des équilibres technologiques adoptés, et du fait de l'hétéronomie induite par le recours à un éditeur XML commercial, les efforts de développements se sont concentrés quasiment exclusivement sur l'interface publique de Calames, et les équipes de l'ABES se sont cantonnées aux mises à jour nécessitées par les évolutions Microsoft ou induites par les modifications des bonnes pratiques de catalogage (une attention plus particulière ayant été donnée à la création et à l'amélioration d'exports XSLT).

L'emploi du plugin XMAX, qui repose sur des technologies Microsoft (ActiveX), suppose un usage exclusif du navigateur Internet Explorer (versions 7/8 et suivantes). La gestion des droits d'accès mise en place par l'ABES fait par ailleurs appel à un composant Net Framework, autre technologie propre aux systèmes d'exploitation Windows. Cette inféodation à l'environnement technologique Microsoft, qui peut s'inscrire en porte-à-faux avec les politiques informatiques de certains établissements et nécessiter l'installation de postes dédiés, représente aussi un handicap certain dans la maintenance et dans l'évolutivité de l'outil de catalogage : bugs parfois peu évidents à expliquer et à résoudre, course sans fin au rattrapage et à la correction des codes suite aux mises à jour Microsoft – dont il n'est pas exclu qu'ils puissent un jour être fatals à l'outil (marginalisation de la technologie ActiveX notamment, et impossibilité de faire fonctionner l'outil sur des versions 64 bits).

Les contraintes propres au fonctionnement de l'outil de catalogage (dépendance Microsoft, compromis techniques remontant aux développements de 2006-2008) se conjuguent avec des facteurs organisationnels (les catalogueurs n'étant souvent pas administrateurs de leurs postes de travail ni maîtres des règles de sécurité de leurs réseaux) et/ou avec des performances clients et réseaux plus ou moins bonnes, et/ou avec un manque de pratique régulière¹², favorisant des situations de travail pénibles voire démotivantes : pertes de données non enregistrées (malgré le filet de la récupération des fichiers temporaires en cours d'édition)¹³, freezes du navigateur IE¹⁴, les temps de réponse à l'enregistrement des

d'inventaires trop lourds fragmentés en plusieurs instances, la liaison de fichiers permet aussi de faire publiquement abstraction de l'existence de plusieurs inventaires pour un même fonds ou collection (ex. : inventaires CGM et leurs suppléments).

¹¹ Export XSLT « visio_controle », dont la 1^{ère} version a été développée en 2010 par Enrico Cima.

¹² Il est patent qu'une plus grande familiarité avec les outils permet de mettre en place des stratégies de contournement et de s'accommoder des lenteurs, lourdeurs et risques qui leurs sont propres. Cf. la mise en place parfois difficile de la cinématique Calames-IdRef en 2011 : dans certains cas, les problèmes rencontrés ne tenaient pas tant à des blocages techniques qu'à une incompréhension des inerties d'affichage de l'éditeur et/ou des enjeux spécifiquement liés à l'indexation.

¹³ Le piège de la touche « retour arrière », qui éjecte l'utilisateur dont le dernier clic souris a porté sur une zone extérieure aux trois volets de l'outil, fait d'ailleurs partie des points d'alerte des formations Calames.

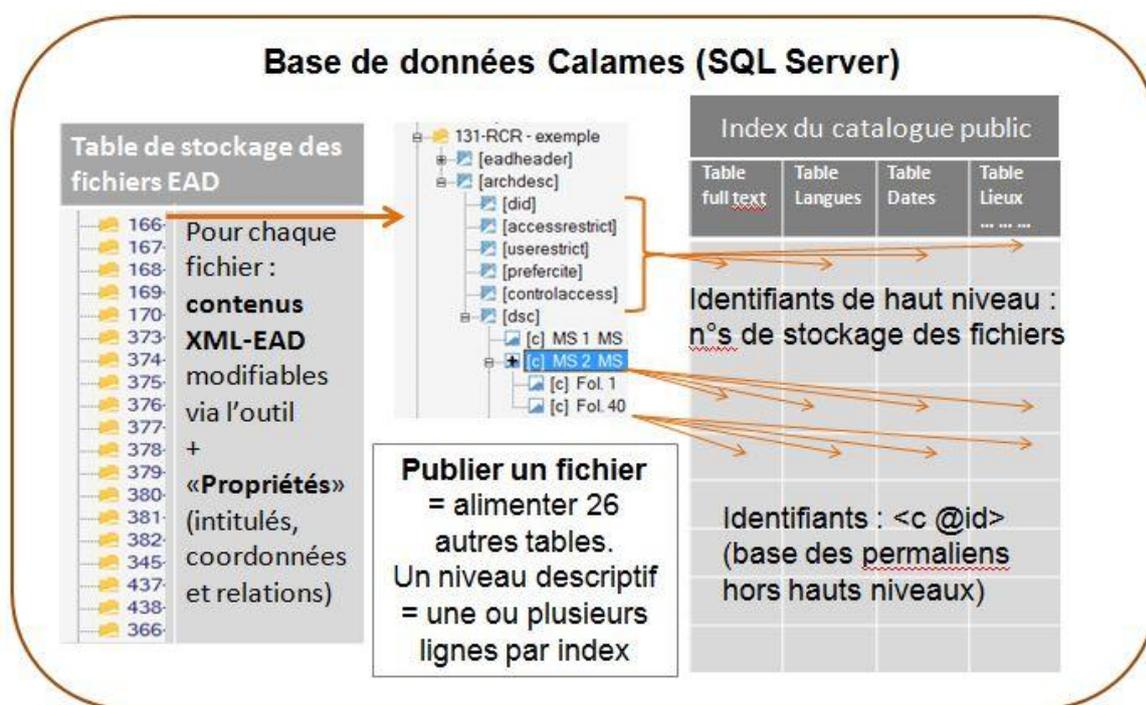
¹⁴ Message d'alerte « Internet Explorer ne répond plus ».

données¹⁵, perte d'autonomie dans les exports (visio_controle, Marc...) du fait des trop grandes masses de données à traiter en ligne¹⁶.

[Voir en annexe 1 les chiffres relatifs à l'activité de catalogage depuis 2008.]

Calames, une interface publique

Les données stockées nativement en XML-EAD sur les serveurs de production peuvent être publiées sur demande d'utilisateurs disposant de droits en édition. Les identifiants attribués par l'outil de catalogage à chaque niveau descriptif (attributs ID de composants <c>, et clés de fichiers EAD pour l'identification des hauts niveaux d'instruments de recherche) sont alors exploités comme bases des permaliens¹⁷, et partant, de la citabilité et d'une partie de la réexploitation des données de Calames. Un processus de publication est une action relativement lourde qui n'est censée être entreprise que d'une manière réfléchie¹⁸.



¹⁵ Ces temps d'attente sont dus tantôt à l'importance des trames XML transitant entre le poste de travail et le CINES, tantôt au vieillissement ou aux limites du poste de travail et/ou des OS Windows. Les établissements du réseau connaissent des situations et des degrés d'insatisfaction très contrastés à cet égard. Les moissonnages systématiques de données par des robots peuvent aussi ponctuellement engendrer des problèmes de performance.

¹⁶ En-dehors, ou en plus, des lenteurs de l'application soulignées par divers établissements, les exports non-natifs (html, visio_controle, marc...) deviennent rapidement impossibles au-delà d'un poids de 1 à 1,5 Mo (soit nettement moins que la taille maximale recommandée de 3 Mo). L'ABES s'est jusqu'à présent refusée à abaisser le poids maximal autorisé des instances pour ce seul paramètre dû aux limites d'un traitement en ligne : l'équipe Calames propose comme palliatif la fourniture des fichiers d'export après demande via son guichet d'assistance et recours en interne à une transformation hors réseau. Une solution d'export côté client pourrait également être envisagée.

¹⁷ Les identifiants attribués par l'outil Calames sont uniques à l'échelle de la base et sont un matériau fiable pour constituer des URL ; ils ne constituent toutefois pas des identifiants pérennes *stricto sensu*.

¹⁸ Au cours d'un processus de publication, les index simple et avancés du catalogue sont mis à jour par alimentation des tables d'une base SQL Server, sur le principe « une unité descriptive = un (ou plusieurs) enregistrement(s) », et en tenant compte de l'héritage le cas échéant (index simple, des points d'accès, des dates, des langues, des numérisations).

Le catalogue public Calames, qui repose sur une architecture Ajax¹⁹, interroge directement cette base de données à travers des web services spécifiques. Calames propose une navigation horizontale et verticale au sein des inventaires²⁰, une rationalisation des lots de résultats par « infanticide » (suppression des unités descendantes d'autres unités au sein d'un lot²¹) et des filtres et facettes à plusieurs niveaux (établissements, types de documents...) reposant sur les index. L'optimisation des résultats dans l'interface publique dépend d'un équilibre entre dispositifs techniques et choix d'encodage (qualité des données et respect des bonnes pratiques)²². A partir de sa mise en service en déc. 2007, l'interface publique a été enrichie de nombreuses fonctionnalités (par ordre chronologique) :

- association automatisée de vignettes cliquables dans la boîte à outils, d'abord pour relier Calames à Liber Floridus (base d'enluminures hébergée au CINES) et étendue au signalement des numérisations déposées dans le cadre du service « Calames Plus » (*cf. infra*). Ces vignettes cliquables apparaissent semblables à celles générées via l'encodage EAD pour les numérisations non hébergées par l'ABES (BVMM, etc.)
- interfaces personnalisées (habillage graphique de la base Calames restreinte à un RCR) pour prendre en considération les exigences de visibilité institutionnelle des établissements producteurs²³
- version anglaise de l'interface publique (nov. 2010)
- services « Calames Plus » (juin 2012) divisés en deux volets : [commentaires](#) (pour tout internaute créant un compte Calames)²⁴, et visionneuse d'[images associées](#) aux notices (après dépôt et tuilage sur serveurs de l'ABES)²⁵
- nouveaux index (documents numérisés, typologies documentaires, techniques) mis en service en mai 2013
- services complémentaires au sein de la boîte à outils : bouton « contacter l'établissement », mise à disposition d'exports PDF.

¹⁹ Ajax (Asynchronous JavaScript and XML) est une méthode de développement des sites internet, permettant d'effectuer des requêtes à l'intérieur d'une page web sans recharger la page (gain considérable en termes d'interactivité et d'ergonomie), en se fondant sur les technologies HTML et CSS pour la présentation des pages, DOM pour la représentation des données, et XMLHttpRequest pour la manipulation des requêtes et réponses.

²⁰ Feuilletage clarifié par la mise en place d'arborescences de fichiers (pyramidage EAD – et totalement en EAD – à comparer avec les « cadres de classement » ou arborescences de dossiers qui commandent l'accès aux instruments de recherche dans les interfaces publique Pleade), ainsi que par la présence et/ou le rappel d'informations parentes dans l'affichage détaillé de tout niveau descriptif.

²¹ Il est à noter que pour des raisons de performance, cette fonctionnalité est limitée aux 5000 premiers résultats d'une recherche. Ce seuil aurait lieu d'être rehaussé au vu non seulement de l'accroissement de la quantité de données publiées, mais aussi d'une profondeur des instruments de recherche souvent plus grande dans le cas des fonds nouvellement signalés.

²² Penser par exemple aux dérogations à l'héritage XML prévues par Calames dans l'indexation (et par voie de conséquence dans l'encodage) des langues et des dates.

²³ En avril 2014, huit établissements Calames bénéficient d'une interface personnalisée : BSG, Institut de France, BDIC, BLJD, Muséum, Musée Curie, Académie de Médecine, Mazarine, CNAM. A noter : la principale faille du dispositif de customisation réside dans l'impossibilité de restreindre les rebonds des points d'accès aux seules données du RCR concerné.

²⁴ Le comité de suivi et le GT Calames ont plusieurs fois fait part de leurs doutes quant au volet commentaires, authentique échec du fait notamment de la conception en « cul-de-sac » des comptes Calames, déconnectés de tous les réseaux sociaux.

²⁵ Le dépôt d'images ne peut être effectué que par des professionnels disposant d'un compte Calames d'établissement, et après préparation des lots à déposer (images jpeg ou gif convenablement nommées) sous forme de fichiers zippés de 100 Mo maximum. La structure arborescente d'une description peut être transposée à l'aide du « cumul d'images » (les images déposées sous des niveaux enfants sont additionnées et rendues toutes accessibles au niveau parent). Le logiciel Zoomify (version gratuite) est utilisé à la fois pour tuiler les images déposées, et comme outil de visionnage. Ce service à portée et capacité limitées reste en-deçà des attentes d'un certain nombre d'établissements et d'utilisateurs quant à la mise en place d'une véritable bibliothèque numérique mutualisée (*cf. infra*).

Bien que les choix techniques opérés en 2007 n'aient pas tous fait la preuve d'une évolutivité optimale à l'occasion des développements postérieurs à la phase projet²⁶, l'interface publique de Calames a jusqu'à présent été considérée comme une réussite en termes d'ergonomie et de performances, voire comme une vitrine pour l'ABES. Certains développements envisagés un temps n'ont à ce jour pas reçu de traduction ou n'ont pas été jugés prioritaires (paniers de notices et exports de références bibliographiques, prise en compte de l'attribut @audience...).

Quant aux usagers du catalogue public, les informations à leur sujet ne sont que partielles. Les niveaux de consultation de Calames sont manifestement comparables à ceux de BnF Archives et manuscrits. Les usages (semi-)professionnels se conjuguent à une part non négligeable de sérendipité, que permet l'exposition des composants publiés auprès des moteurs de recherche (protocole sitemap).

[Voir en annexe 4 quelques chiffres et commentaires relatifs à la consultation de l'interface publique.]

Calames, outil collectif et collaboratif

Une des caractéristiques les plus originales de Calames est d'être né d'une collaboration étroite entre l'ABES et un ensemble d'établissements particulièrement impliqués dans la question du signalement patrimonial²⁷. Ce principe de collégialité se perpétue notamment par la tenue régulière de réunions du groupe de travail Calames²⁸, rassemblant les équipes dédiées de l'ABES et des établissements qui, par l'ampleur ou la nature spécifique de leurs collections, leur implication dans des projets nationaux de recherche ou de valorisation, l'expertise de leur représentant, fournissent à l'ABES des renseignements sur les pratiques et les attentes des utilisateurs et la conseillent sur les évolutions souhaitables.

Calames, qui ne saurait être du fait de la nature unique des documents signalés un catalogue partagé²⁹, n'est ni seulement un outil commun fourni clés en main à différents établissements, ni seulement une juxtaposition de catalogues interrogeables simultanément dans une même interface, mais bien la mise en réseau des compétences et des expériences diverses des établissements producteurs. La mise à disposition de documentations et d'outils communs, l'homogénéisation des pratiques d'encodage (gage et condition de la qualité des recherches, assurée par le Manuel Calames qui respecte et précise les bonnes pratiques EAD dans les bibliothèques françaises), une gestion de la mutualisation souvent inspirée du Sudoc (conventionnement et facturation de licences, liste de diffusion pour informations ABES et échanges entre catalogueurs, formations de déploiements et formations-relais ou en ligne)³⁰, sans oublier l'existence d'un dispositif de soutien financier aux conversions rétrospectives³¹, concourent à faire de Calames le seul véritable réseau³² français de bibliothèques productrices de données en EAD.

²⁶ Par exemple : choix de la librairie javascript « Prototype », en déclin aujourd'hui (contrairement à JQuery, Dart... plus adaptables et polyvalentes).

²⁷ Ces établissements sont notamment prêts à autoriser leur correspondant Calames à consacrer une part non négligeable de son temps de travail à l'expertise de dossiers communs.

²⁸ Composition actuelle : <http://www.calames.abes.fr/pub/groupe.aspx>

²⁹ Si chaque établissement producteur est le seul à avoir accès à ses propres données en mode catalogue (modifications, suppressions, enrichissements), tous les établissements du réseau ont accès en lecture seule à l'ensemble des instances EAD présentes dans la base de production (publiées ou non), afin de leur permettre de s'inspirer d'exemples d'encodage extérieurs.

³⁰ Les échelles, enjeux, schémas techniques ne sont pas toujours comparables à ceux du Sudoc, loin s'en faut, ce qui peut parfois entraîner des inadéquations. La liste de diffusion Calames est ainsi notablement sous-employée par le réseau : elle sert *de fait* principalement de canal d'information de l'ABES vers les établissements.

³¹ De 2009 à 2014, un montant de 629.179 euros a été distribué à 21 établissements Calames dans le cadre de la mission rétroconversion, dont la responsabilité est passée en 2010 de la SDBIS à l'ABES. On peut estimer que les activités de catalogage dans l'outil Calames ont été soutenues à hauteur de 30 à 40% par ces sommes.

³² En 2014, 49 établissements conventionnés correspondant à 57 RCR avant déploiement à la fin de l'année d'un 8^e cercle.

La cohérence du modèle Calames pourrait être questionnée au regard de l'hétérogénéité du réseau dans son activité de création et d'amélioration des métadonnées EAD³³. Le réseau Calames est nettement marqué par la prégnance de quelques établissements très actifs, souvent impliqués dès l'origine du projet, et très souvent parisiens ou franciliens. A contrario, certains déploiements se sont jusqu'alors traduits par une production de nouvelles données très faible voire nulle, et/ou par un usage fortement raréfié de l'outil une fois les traitements rétrospectifs achevés³⁴.

De plus, le caractère peu contraignant d'un système auquel l'adhésion n'est conditionnée qu'à l'appartenance à un périmètre d'ordre essentiellement institutionnel et à des typologies documentaires qui se sont diversifiées, et qui fait reposer l'essentiel du respect des règles communes sur l'autodiscipline des participants (y compris dans le cadre des opérations de conversions rétrospectives financées par l'ABES), est porteur en interne d'un risque de divergence des pratiques potentiellement préjudiciable à la cohérence même des données ou à leur exploitation, et en externe, d'un risque de fracture avec d'autres acteurs majeurs de la recherche et de la production de métadonnées EAD.

³³ Au cours des cinq premières années d'utilisation de l'outil de catalogage de Calames, les deux tiers des 443000 composants <c> créés via cet outil l'ont été par cinq établissements seulement : le Muséum (24%), l'Institut de France (15%), l'Académie de Médecine, la BDIC et la BLJ Doucet (10% chacun).

³⁴ L'écart peut être très important entre d'une part des établissements rodés à l'outil et très habitués aux modes de fonctionnement du réseau, et d'autres part des établissements aux usages ponctuels et qui ne s'approprient Calames que très lentement et/ou très partiellement. Une enquête menée durant l'hiver 2012 auprès de neuf établissements alors déployés, mais non productifs, avait permis de repérer plusieurs causes d'inactivité dans Calames : manque de temps et de moyens humains (dans 2/3 des cas), concurrence avec d'autres projets (y compris description de fonds imprimés dans le Sudoc), mais aussi problèmes d'accès à l'interface de catalogage (proxys, mises à jour Windows) et besoins en formation et/ou d'échanges accrus avec d'autres catalogueurs Calames.

Enjeux et perspectives

Aussi nombreuses les pistes d'évolutions des outils et du modèle de réseau Calames soient-elles, il est essentiel de prendre en considération la petite taille de l'équipe Calames à l'ABES (environ 3 ETP consacrés à tous les aspects de l'application, de la maintenance et des développements jusqu'aux formations et au suivi des rétroconversions). L'implication parfois très active d'une partie du réseau (dans le cadre du GT Calames, des formations-relais, voire dans la veille sur les questions d'outils et de métadonnées), et/ou la conclusion de nouveaux partenariats sont indispensables à la perpétuation, et plus encore à tout développement ou reconfiguration de Calames.

Faire évoluer le réseau

Calames ne peut ni ne doit cultiver sa spécificité purement conjoncturelle mais se concevoir et se développer en lien avec les autres catalogues ou réseaux recensant des collections similaires ou complémentaires, tant au niveau national (CCFr, BAM, BVMM, Archives de France et archives départementales) qu'au niveau international (CERL, APEnet, etc.). La poursuite et le renforcement des collaborations déjà engagées ne doivent pas dispenser d'une réflexion sur les moyens d'en mettre en œuvre d'autres. Les synergies internes doivent être favorisées pour dépasser le simple signalement des collections de chacun en facilitant (mise à disposition d'outils, soutien scientifique et financier) les partenariats et les valorisations collaboratives (constitution et traitement de corpus par sujet, auteur, provenance ou autre).

Le cadre actuel des déploiements dans Calames, largement inspiré du « modèle Sudoc », fondé sur l'usage (facturé) d'un outil de production principal et partagé par tous³⁵ ainsi que sur la référence à des bonnes pratiques communes, a lieu d'être repensé pour prendre en compte, sans sacrifier la nécessaire uniformisation des pratiques³⁶, la diversité des choix de production et de signalement préexistant à l'adhésion au réseau³⁷, ainsi que les besoins de valorisation propres à chaque établissement sans renoncer à une visibilité collective nationale et internationale³⁸.

D'un point de vue institutionnel, et en vue de faire de Calames la « base patrimoniale de l'Enseignement supérieur et de la Recherche », il est opportun de faire évoluer le périmètre actuel du réseau Calames en direction des producteurs ou gestionnaires d'archives du monde de l'ESR (laboratoires, centres de recherche, services d'archives, etc.), quel qu'en soit le statut³⁹. Au sein même du périmètre relevant du MESR, il semble important d'amener les institutions patrimoniales de l'Enseignement supérieur qui n'ont

³⁵ Le comité de suivi Calames a fermement affirmé, à plusieurs reprises, que Calames ne pouvait être considéré seulement comme une « vitrine publique » diffusant des données produites à l'extérieur du réseau : le respect des bonnes pratiques étant l'un des gages offerts par l'usage de l'outil Calames, et les services associés à l'application (auxquels on peut aller jusqu'à inclure la possibilité d'un soutien financier via dossiers de rétroconversion) n'étant pas suffisamment développés pour justifier un conventionnement sans inclusion d'au moins une licence XMetal. Ce positionnement distingue tout particulièrement le réseau Calames des établissements hors ESR signalant leurs inventaires EAD via le CCFr Manuscrits, et a amené plusieurs établissements à devoir considérer l'outil professionnel Calames comme le nouveau pivot de leur production en EAD – ou à tout le moins, à l'intégrer comme environnement de traitement secondaire de leurs données.

³⁶ Ce qui supposerait de vérifier que l'ensemble des fichiers EAD – tant importés que produits – respectent strictement les préconisations du Manuel Calames. Le dispositif actuel ne permet pas un tel niveau de contrôle.

³⁷ Usage d'éditeurs XML tels qu'oXygen ou ICA-Atom, de logiciels de gestion d'archives type Arkheia, voire même (très à la marge) intégration de notices de manuscrits en Marc dans le Sudoc...

³⁸ Certaines fonctions strictement locales (p.ex. visualisation de reproductions numériques de documents sous droits) sont assurées par certaines bases Pleade, mais l'existence de ces systèmes locaux au sein du réseau Calames (BLJD, Observatoire, Cirdoc, MSH...) entre au moins partiellement en redondance avec l'offre de customisation des interfaces de recherches.

³⁹ On peut notamment penser aux organismes de recherches dépendant d'autres secteurs ministériels (Santé, Industrie, etc.), ainsi qu'aux sociétés savantes et associations disposant de fonds d'archives intéressant la recherche. Un contre-exemple récent pourrait être cité : celui de l'École polytechnique, assimilée au périmètre ESR par héritage des rétroconversions CGM et Palme, qui a marqué dernièrement une préférence pour un signalement via le CCFr Manuscrits.

pas inscrit dans leurs priorités le traitement de leurs fonds d'archives ou ont opté pour d'autres choix d'encodage et de signalement, à s'investir pleinement dans le réseau⁴⁰ et à utiliser l'outil dans toutes ses composantes.

Une attention renforcée doit également être portée envers les établissements dont les moyens ne sont pas en adéquation avec l'importance des fonds à signaler, en développant des dispositifs d'accompagnement et d'incitation à produire dans Calames dans le respect des pratiques communes. Face aux risques de mauvaises pratiques EAD et de méconnaissance des outils, il convient de mettre à disposition du réseau des moyens évolutifs et efficaces de contrôle des données (jusqu'alors indissociables de l'usage de l'outil de catalogage) et de leur suivi⁴¹, ainsi qu'une offre de formations adaptée et adaptable⁴² et un renforcement des dispositifs de conseil⁴³. Sans remettre en cause le principe d'autodiscipline et d'autonomie de chaque établissement, l'une des orientations actuelles de l'ABES est de mettre en place un contrôle qualité dans le cadre du suivi des projets de rétroconversion qu'elle cofinance⁴⁴.

Remédier aux limites de l'outil induites par les choix technologiques initiaux

Il convient de réfléchir aux moyens à mettre en œuvre pour réduire substantiellement, voire faire disparaître la « dépendance » de l'outil à des évolutions progiciels subies du fait des choix techniques initiaux (cf. *supra*)⁴⁵.

Le modèle d'utilisation en ligne de l'outil, qui pose parfois de sérieux problèmes de performances, doit être repensé à l'aune des habitudes de travail constatées ou anticipées. Si certaines fonctionnalités ne peuvent être exécutées que sur les serveurs ABES (publication totale ou partielle notamment) et si d'autres peuvent être maintenues en ligne (consultation en lecture seule de l'ensemble des instances EAD, gestion des fichiers), la plupart sinon la totalité des fonctionnalités d'édition / encodage pourraient gagner à être prises en charge par un outil d'édition client complet permettant le téléchargement temporaire d'un fichier XML le temps de l'encodage et son « renvoi » / copie sur les serveurs de l'ABES. Une telle option acterait le caractère accessoire du « versioning » (pouvoir éditer simultanément et à plusieurs une même instance), exigence fonctionnelle qui détermina des choix techniques et des développements parfois coûteux, alors même qu'il est fort rare, en pratique, que plus de deux catalogueurs d'un même groupe de droits aient à intervenir simultanément sur un même fichier EAD.

Assurer la cohérence, la mise à niveau et la qualité des données

Une part importante – quoique en proportion sans cesse décroissante – des données de Calames sont issues de la conversion rétrospectives du CGM, ensemble de catalogues papier dont la rédaction s'était échelonnée sur un siècle et demi. Malgré le travail réalisé alors par le groupe des établissements ESR pilotes sur la totalité des données, il n'a été possible ni d'harmoniser totalement la structuration des

⁴⁰ Ce qui suppose, au-delà des avantages les plus évidents (plate-forme de diffusion plus large, recours aux appels à financements pour rétroconversion, sécurisation des données, etc.), le respect des règles communes d'encodage.

⁴¹ Actuellement, aucun historique des publications dans Calames (de type wiki) n'existe ; on ne peut connaître, pour un niveau descriptif donné, que la date de sa dernière (re-)publication. L'ABES pourrait conserver une trace précise des nouvelles publications de données, constituant la base d'un véritable tableau de bord du contrôle qualité.

⁴² Les formations-relais Calames ont été conçues à partir de 2012 pour qu'un tronc commun soit modulé en fonction de deux types de publics : débutants ou utilisateurs confirmés.

⁴³ Que ce soit sur la liste calames@liste.abes.fr ou directement auprès de l'équipe Calames, les questions de métadonnées et les demandes spontanées de conseils d'encodage sont toujours les bienvenues, et devraient même, si cela n'engageait pas des questions de ressources humaines, être plus ouvertement encouragées.

⁴⁴ Un enjeu constant pour l'ABES est de trouver un juste équilibre entre un non-interventionnisme contraire à l'esprit de réseau et un soutien aux rétroconversions qui représenterait une part trop importante de l'activité générale dans Calames.

⁴⁵ Il serait opportun de procéder *a minima* à une mise à jour de la version du plugin XMAX (de la version 5.0 à la 7.0), voire, dans le cas d'une refonte globale de l'interface publique, à celle de certaines librairies Javascript.

descriptions⁴⁶, ni d'assurer un lien vers des référentiels d'autorité⁴⁷. La mise à niveau de ces descriptions, et notamment en ce qui concerne la normalisation et le liage des points d'accès, doit impérativement être pensée en parallèle avec la production de nouvelles données.

Depuis lors, malgré l'adoption au niveau national de recommandations pour le signalement des manuscrits et archives (DeMArch), de bonnes pratiques conformes à ISAD(G) et leur déclinaison dans le Manuel de catalogage Calames, les données produites dans la base, parfois issues de rétroconversions locales, sont encore trop hétérogènes et non conformes à ces bonnes pratiques. La mise en place d'un schéma conforme aux règles du Manuel et auquel serait conditionnée la publication des données, tout comme l'encouragement à la reprise de données non conformes, permettrait de renforcer le contrôle qualité en conformité avec des règles nationales.

L'enrichissement des descriptions en regard des standards scientifiques actuels et des exigences des utilisateurs doit être encouragé. Pour assurer la crédibilité scientifique de Calames et éviter la multiplication de descriptions à la fois concurrentes et divergentes, le résultat des recherches menées depuis la rédaction initiale des notices doit être au maximum intégré⁴⁸. Cette valorisation des descriptions, notamment bibliographique, étant ou devant être le reflet d'une véritable insertion dans le monde de la recherche au-delà d'un simple signalement de collections.

S'adapter à l'évolution du contexte normatif et réglementaire

L'emploi de l'EAD pour le signalement dans Calames ne saurait être remis en cause, alors même que le choix de ce standard pour décrire les manuscrits et archives est partagé par tous les acteurs nationaux (Archives de France, BnF, CCFr) et que les praticiens sont de mieux en mieux formés. Néanmoins, les évolutions en cours de la DTD risquent d'avoir un impact important sur les outils, les préconfigurations du nouveau schéma EAD 3 semblant parfois peu compatibles avec l'existant. Sans préjuger de l'avis du groupe AFNOR CG46/CN357/GE3 qui suit ce dossier au niveau national⁴⁹, ni des décisions qui seront prises par les différents acteurs concernés, il est primordial que Calames puisse, au minimum, rester compatible en ce qui concerne les recherches fédérées avec les autres catalogues nationaux (*cf. infra*).

Développer l'interopérabilité

Les réflexions en cours sur les modèles de données, qu'ils soient bibliographiques, muséaux ou archivistiques (FRBRoo, modèle CIDOC-CRM), doivent également sous-tendre les évolutions de Calames. La diversité de nature (textes manuscrits, imprimés, enregistrements analogiques ou numériques, objets manufacturés ou non, etc.), la spécificité (original, brouillon, copie, reproduction, etc.) et l'importance des données concernées qui sont parfois des sources uniques d'information, participent en effet pleinement à ces modèles et doivent pouvoir s'intégrer dans les recherches du web sémantique.

Le caractère « généraliste » de Calames qui doit permettre, conformément au périmètre qui lui a été assigné (*cf. supra*), la description de fonds archivistiques, de manuscrits mais également d'imprimés, d'objets, etc., suppose une attention particulière apportée aux liens qui peuvent être faits avec des catalogues ou des bases spécialisées permettant une description plus fine de certains composants de nature spécifique

⁴⁶ Des informations de même type peuvent, selon les fichiers, être encodées de manière différente.

⁴⁷ Dans le meilleur des cas, ont été saisies des formes normalisées et si possible conformes aux formes retenues proposées par les référentiels alors disponibles de la Bibliothèque nationale de France ou du Sudoc.

⁴⁸ Pour permettre un suivi documentaire, on peut envisager de « versionner » les descriptions selon les modifications apportées (modifications dont la source doit être mentionnée).

⁴⁹ Dans le cadre d'une réflexion nationale (bibliothèques & services d'archives) sur ces évolutions, il paraîtrait intéressant qu'une étude chiffre les impacts techniques (perte d'information notamment mais également refonte des bonnes pratiques et des contrôles en dépendant) et financier (conversions rétrospectives à prévoir) sur la masse de données EAD existantes, au-delà de ceux de la mise en place d'un nouvel outil compatible avec cette version.

(imprimés, objets, etc.) ou de certains éléments descriptifs (reliures, provenances, décor, etc.). La gestion de ces liens, leur réciprocité, la possibilité de les utiliser pour enrichir des index et / ou des affichages sont autant de points à étudier soigneusement.

Extension particulière du point précédent, la gestion des autorités des formes indexées quelle que soit leur nature (personnes physiques, collectivités, familles, titres d'œuvres, noms géographiques, sujets, typologies, autres) ou leur rôle, apparaît cruciale tant dans l'immédiat qu'en vue d'une intégration des données dans un modèle FRBRisé⁵⁰ ou dans les outils utilisant les technologies du web de données. Les efforts fournis dans ce domaine doivent être poursuivis et s'intéresser, dans une logique d'échange avec le monde des archives, aux travaux sur le format EAC-CPF, comme aux réflexions en cours sur les référentiels (hypothèse d'un réservoir national partagé, VIAF, Geonames, etc.)⁵¹.

Favoriser l'exploitation des données

La multiplication des projets de recherche et / ou de valorisation – humanités numériques telles que bibliothèques numériques, éditions électroniques, etc. – sur les fonds faisant l'objet de descriptions dans Calames implique une réflexion approfondie sur l'utilisation de ces données descriptives comme source unique et fiable des métadonnées de ces projets. Cela passe, d'une part par une analyse des données potentiellement concernées et le développement d'exports simples dans des formats ou standards aussi divers que Dublin Core, TEI, RDF... ; d'autre part, par le fait de rendre possible de manière simple, par moissonnage, la constitution de bases spécifiques correspondant à un corpus donné (typologie documentaire, provenance, etc.) ; faute de quoi, on assisterait, pour un même objet, à une multiplication de signalements différents dans la masse desquels celui donné par Calames apparaîtrait noyé et sans crédibilité scientifique plus manifeste.

L'une des problématiques propres aux bases EAD est qu'elles souffrent d'une forme d'isolationnisme subi, du fait de la spécificité et de la complexité de leurs fichiers arborescents⁵² (cf. *infra* § « numérisation »). L'adéquation des choix d'exports à leurs divers cadres de réexploitation est une souplesse cruciale que se doit de développer Calames. Le moissonnage *ad hoc* mis en place par le CCFr Manuscrits à partir de février 2013 peut être cité, sinon en modèle, du moins en exemple : pour prendre en compte les données publiées dans Calames dans toute leur richesse et leur actualité, le CCFr réexploite ses web services natifs (non standards, bien que générant des réponses qui encapsulent pour chaque niveau descriptif un EAD intact).

En plus des exports de données, il faut également prendre en considération les besoins des établissements producteurs en matière d'articulation de Calames avec leur environnement local, qu'il s'agisse d'interrogations (portails, sites, etc.) et / ou de communication (SIGB) et /ou de gestion d'informations à usages internes. Parmi les demandes récurrentes d'établissements figurent ainsi divers besoins de gestion des collections (communication, statistiques, GED, dossiers d'acquisition, récolements, informations internes voire confidentielles⁵³...) : sans aller jusqu'à justifier la mise en place de services relevant de

⁵⁰ La question des titres d'œuvres notamment, points d'accès dont les emplois sont sujets à une grande variabilité dans Calames, aura tout lieu d'être envisagée à l'occasion des travaux en cours sur l'implémentation des FRBR.

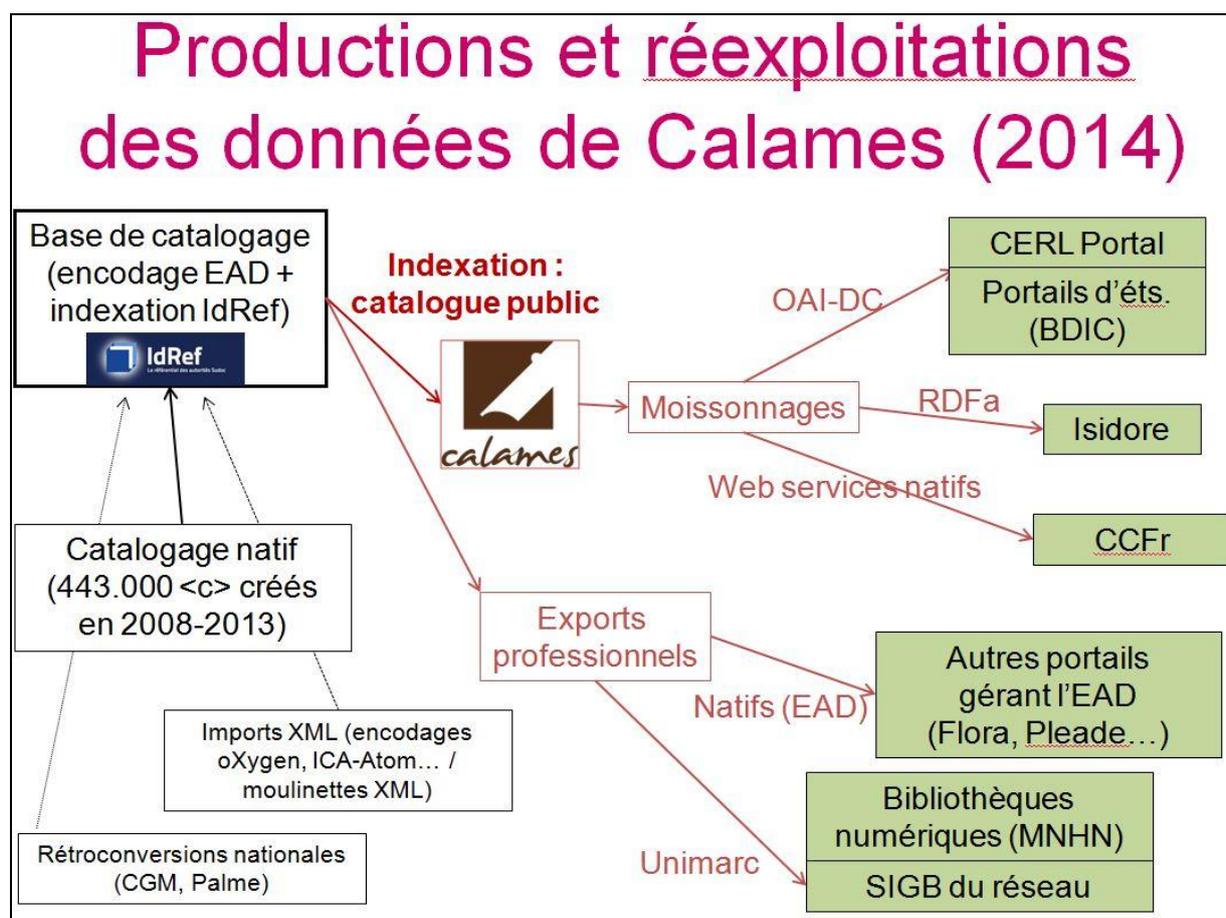
⁵¹ Mentionnons également le cas particulier du thésaurus *ad hoc* réalisé et maintenu par le GT Calames pour la normalisation des points d'accès « genre, forme et fonction », s'inspirant de plusieurs référentiels existants (Rameau, thésaurus archives locales du SIAF, RBMS Genre Terms).

⁵² Cette structure arborescente pourrait être exploitée via le moissonnage de Calames en RDFa par des liens entre niveaux descriptifs ; ce que ne fait toutefois pas le portail Isidore, seul outil jusqu'alors à exploiter cette syntaxe des données publiées.

⁵³ Le cœur du problème n'est pas dans l'extension des usages ou dans le dévoilement des instruments de recherche. L'ABES maintient son choix de ne pas prendre en compte l'attribut @audience="internal" : il serait trop coûteux en re-développements d'exclure certains éléments ainsi qualifiés non seulement à l'affichage, mais aussi à l'indexation. Les commentaires XML, sans garantir une parfaite confidentialité, sont semble-t-il un moyen terme acceptable pour inclure des données qui n'ont pas lieu d'être rendues interrogeables ni visibles.

Systèmes d'Information Archivistique (SIA), cet enjeu ne saurait être relevé par de simples adaptations dans l'encodage des instruments de recherche. La réponse à ces attentes doit être envisagée en lien avec la réflexion menée à l'ABES sur le SGBm et / ou des développements des interfaces locales personnalisées au-delà des seules adaptations graphiques.

De même, les attentes, besoins ou habitudes des chercheurs en termes de référencement et de récupération de données doivent, malgré les difficultés inhérentes à l'identification des données pertinentes, être prises en compte en proposant des exports dans des formats reconnus et utilisés par le monde de la recherche et de l'édition. Ces (ré)utilisations des données de Calames, lui assurant un rôle de « colonne vertébrale » des différents projets liés aux fonds qui y sont décrits, supposent une réflexion parallèle sur le statut juridique de ces données, afin de prendre en compte leur origine sans bloquer leur usage. Inversement, Calames devrait plus nettement faciliter l'intégration ou le signalement de publications concernant les documents signalés.



Rapprocher signalement et numérisation

Si la numérisation n'entre ni dans les missions originelles de l'ABES, ni dans l'extension plus récente de ses prérogatives vers d'autres domaines d'interventions (licences nationales et groupements de commandes, thèses, etc.), l'émergence et le développement croissant de projets locaux ou nationaux de bibliothèques numériques⁵⁴ concernant des fonds et collections qui, par la prégnance d'un contexte de production et d'un besoin de description arborescente, et/ou par le caractère unique des documents qui les composent, rentrent très fréquemment dans le périmètre documentaire de Calames (manuscrits et

⁵⁴ Exemples notables : BVMM mise en service par l'IRHT, Gallica marque blanche (Numistral à la BNUS)...

documents anciens uniques, photographies, archives de la recherche...), porte à repenser une articulation imparfaite entre signalement et numérisation⁵⁵.

Alors même que l'expérience a montré que les spécificités du signalement en EAD (héritage parfois complexe d'informations) rendent intéressant à plus d'un titre l'accolement direct de reproductions numérisées aux instruments de recherche, l'absence (au-delà du dispositif Calames Plus Images, présenté dès sa conception en 2009-2010 comme un service d'appoint⁵⁶) d'une offre crédible de bibliothèque numérique adossée au catalogue collectif patrimonial auquel ils participent, oblige les établissements à multiplier les interfaces publiques pour leurs données patrimoniales, au risque de redondances ou de surcoûts de production et d'archivage.

Pour être pertinente dans l'écosystème actuel de la numérisation⁵⁷, l'offre Calames aurait lieu de s'étoffer d'un certain nombre de fonctions :

- fonctions de gestion des images (traitements initiaux et postérieurs) à renforcer ; règles de nommage des répertoires et de liage images-niveaux descriptifs à affiner ; etc.
- renforcement de l'ergonomie de consultation (feuilletage, téléchargements...) et meilleure synergie dans l'affichage des métadonnées et des substituts numériques
- mise en place d'un entrepôt OAI⁵⁸
- archivage pérenne des données numérisées liées (gestion de métadonnées METS)⁵⁹ ?
- plus secondairement, ou dans un second temps : élargissement aux numérisations d'archives orales ou audiovisuelles (fichiers MP3, MPEG...) et aux archives nativement numériques ?
- prise en compte des transcriptions textuelles et / ou éditions ou études nativement numériques (via le dépôt de PDF dans un premier temps ?).

Le développement rapide des humanités numériques plaide en faveur de la mise en place d'une véritable chaîne de production numérique intégrant signalement, numérisation et éventuelles éditions électroniques des documents patrimoniaux ; la cheville ouvrière pourrait en être un outil de production national en XML respectant différents schémas en fonction des cadres de travail et projets associés, à l'instar des outils TEI actuellement développés dans le cadre de l'Equipex Biblissima⁶⁰.

⁵⁵ Tout en ayant intégré le patrimoine dans son projet d'établissement, l'ABES n'est pas censée offrir de solution complète à l'un des pans majeurs de la production de données en établissements patrimoniaux. Dans le même temps, l'existence de projets de numérisation parallèles à des projets de description est un des critères retenus par le Conseil scientifique de l'ABES pour présider au choix des allocations de cofinancements pour rétroconversions dans Calames (considérant que la numérisation est un gage supplémentaire pour la complétude et la précision de la description, et considérant que l'accès public aux documents s'en trouve facilité et élargi).

⁵⁶ C'est-à-dire : lorsque l'établissement demandeur ne dispose pas de bibliothèque numérique. Le corollaire jamais explicite aurait pu en être que sont prioritairement ciblés les établissements qui ne souhaitent pas multiplier les interfaces publiques pour leurs données patrimoniales. La modestie des chiffres de dépôts (et donc de consultations) d'images déposées depuis juin 2012 n'a rien d'étonnant (voir les statistiques sur Calames images en annexe), bien qu'au regard des déclarations d'intérêt recueillies dans le réseau, il soit plus surprenant que seuls trois établissements soient jusqu'alors passés aux actes (Institut Curie, Bibliothèque Littéraire Jacques Doucet et Bibliothèque Mazarine).

⁵⁷ La possibilité d'un projet de bibliothèque numérique adossée à Calames, et viable dans l'écosystème national (Gallica, BSN5), ne peut se concevoir qu'au travers d'un partenariat avec des établissements déployés dans Calames, qui manifesteraient un véritable intérêt à disposer d'une bibliothèque numérique adossée au catalogue collectif patrimonial auquel ils participent. Faute de quoi l'ABES serait amenée à laisser en l'état le service Calames images, ce qui reviendrait à laisser en déshérence une raison d'être de plus en plus essentielle à toute base patrimoniale. Certaines de ces réflexions pourraient également valoir dans le cas du Sudoc.

⁵⁸ Le moissonnage OAI-PMH de données Dublin Core n'est actuellement possible qu'à la volée, sans aucun set sur les données numérisées notamment. La mise sur pied d'un tel entrepôt serait par ailleurs profitable à l'ensemble des données de Calames, numérisation ou pas (*cf. chap. interopérabilité*).

⁵⁹ En lien possible avec une fonction de mutualisation des procédures de versement au CINES ; à rapprocher des services proposés par la TGIR Huma-Num ?

⁶⁰ Voir <http://www.biblissima-condorcet.fr/fr/projet/presentation>

Conclusion

La présente étude, basée sur l'expertise des membres du groupe de travail Calames, a bénéficié des apports de l'ensemble du réseau et des personnels de l'ABES. Ces analyses et réflexions ont pu être présentées au cours de la journée du réseau Calames, tenue le 27 mai 2014 à l'Institut de France, au cours de laquelle plusieurs lignes directrices ont été mises en évidence :

- Les établissements du réseau Calames connaissent toujours une très grande variété de situations. Un accent tout particulier doit être mis sur la gestion collective des numérisations, formes devenues incontournables des descriptions de documents patrimoniaux, et sur l'avenir de Calames Plus Images. Jusqu'où redéfinir la notion de « solution d'appoint », tout en évitant l'écueil d'un portail documentaire totalisant ? Une étude spécifique de la question et un positionnement des instances décisionnaires (comité de direction de l'ABES, organes consultatifs du réseau Calames) sont attendus dans le courant du second semestre 2014.
- Partant du constat d'une convergence de besoins à la BnF, dans les établissements CGM, et dans le réseau Calames qui a souligné certains des handicaps de son outil d'encodage, l'hypothèse d'un dispositif national d'édition en XML ouvert et « agnostique » en matière de formats sera examinée à l'automne 2014.
- L'élargissement (y compris en direction des données de gestion interne, hors EAD) et l'approfondissement des solutions de réexploitation des données EAD (web de données, nouveaux webservice...) figureront également dans la feuille de route de l'équipe Calames.

La mise en cohérence des actions de l'ABES avec celles de plusieurs acteurs centraux sur ces dossiers clés (TGIR Huma-Num, Equipex Biblissima...) est indispensable : elle devrait trouver, à partir de l'automne, un cadre privilégié dans le dispositif CollEx.

Annexes statistiques

1. Statistiques par établissements : données publiées, créées, indexées (2007-2014)

Etablissement	Année de déploiement	C publiés fin 2007 ⁶¹	C publiés (mi mars 2014) ⁶²	C créés dans l'outil (2008-2013) ⁶³	Nb total de points d'accès	Nb points d'accès liés à IdRef ⁶⁴
Institut de France	2008	40560 (C)	105555	64690	147249	32330
MNHN		8584 (C)	68172	106011	96897	42216
BLJ Doucet		40979 (P)	60543	44245	121647	42321
BDIC		156 (P)	33774	44554	53231	43530
Bib. Mazarine		15791 (C)	24169	18959	28158	4855
BNUS		10900 (C)	23780	12713	40836	8152
BIU Sorbonne		8494 (C)	23662	13023	27441	5233
BSG		8786 (C)	19196	11907	50391	7869
BIU Montpellier		2652 (C)	7498	3653	7595	3573
Académie de Médecine	2009	1781 (C)	42467	45708	60663	21000
BIU Santé (Pharmacie)		137 (C)	5665	5526	1167	461
BIU Santé (Médecine)		1325 (C)	4375	3174	5095	718
Bib. Thiers		1885 (C)	3605	2466	5700	631
Soc. Hist. Prot. Fra.		1592 (C)	1769	177	7557	358
SCD Paris 8		0	5141	5172	2268	2074
EFA Athènes		0	1489	3719	3125	2932
SICD Toulouse		681 (C)	681	0	1164	13
SCD Nancy		369 (C)	406	39	1202	79
EFEO		0	258	587	1551	1286

⁶¹ La base Calames a officiellement vu le jour en ligne le 1^{er} décembre 2007. L'origine des données est indiquée par les lettres C (CGM) ou P (Palme). Ces chiffres ne sont qu'indicatifs : ils correspondent au nombre de <c> des instances CGM et Palme d'origine, et leur total (158241) est supérieur à la quantité de <c> effectivement rendus publics dans Calames au 1^{er} décembre 2007 (env. 144000). Par ailleurs, en raison de la faible qualité des données, ou en l'absence de réponses de la part de certains établissements responsables (sociétés savantes), env. 5000 <c> CGM et env. 6900 <c> Palme n'ont pas encore été publiés, et/ou ne seront jamais publiés.

⁶² Chiffres recueillis le 18 mars 2014

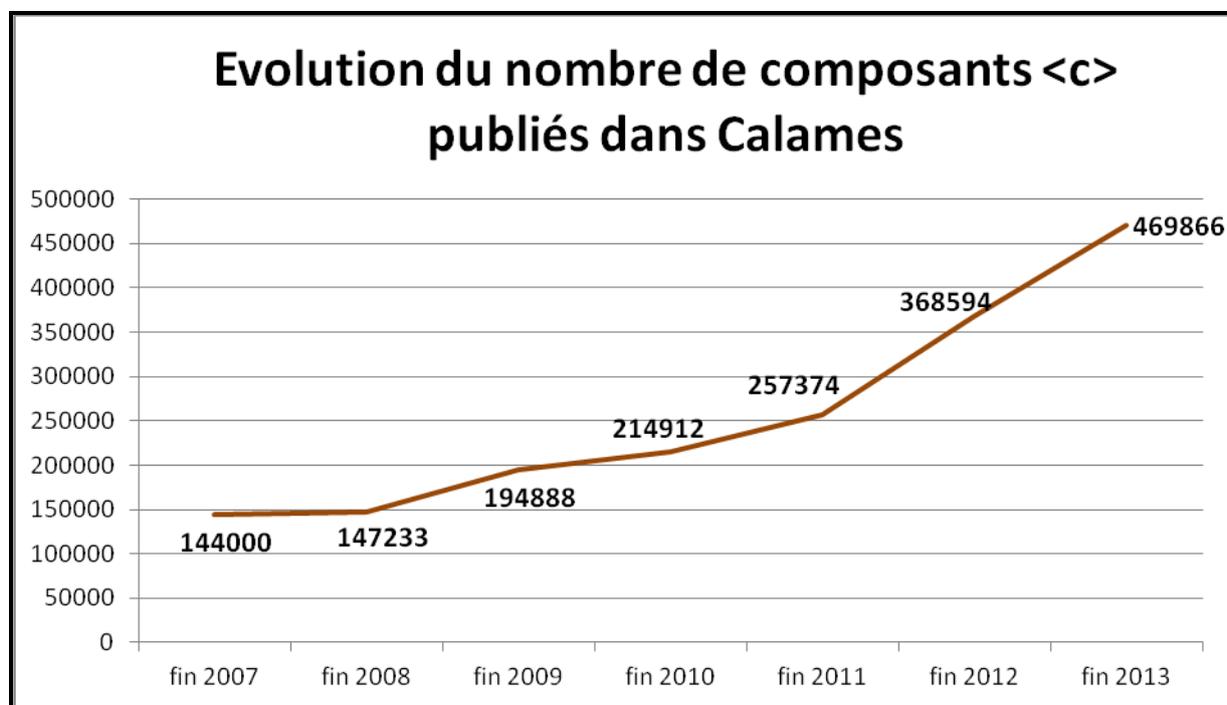
⁶³ Chiffres compilés en janvier 2014. Un différentiel existe en permanence entre le nombre de <c> publiés et le nombre de <c> présents dans la base, dont certains n'ont encore jamais connu aucune publication. Il est délicat, voire spécieux de dénombrer les <c> répondant à ce dernier cas de figure, car ces notices sont plus que les autres sujettes à diverses modifications ou suppressions. A seul titre indicatif, en mars 2014, on pouvait en dénombrer env. 50000 en base de production (base de formation exclue, or certains établissements y produisent aussi de véritables niveaux descriptifs avant de les exporter en base de production). Les établissements les plus représentés dans cette séquence de catalogue particulière étaient alors : INHA (12000), Bibliothèque Mazarine (10700), BDIC (7600), BLJ Doucet (4200), Muséum (3000), SCD Paris 6 (2700), ENS (2300), Académie de Médecine (2000), Bibliothèque de l'Institut (1500), RDD Bordeaux (1500), etc.

⁶⁴ Chiffres recueillis le 18 mars 2014

Etablissement	Année de déploiement	C publiés fin 2007	C publiés (mi mars 2014)	C créés dans l'outil (2008-2013)	Nb total de points d'accès	Nb points d'accès liés à IdRef
Clermont SICD	2009	0	182	227	161	2
BULAC		0	17	41	14	0
SCD Bourgogne		6	6	0	25	1
RDD Bordeaux		216 (C)	3695	5029	1214	1024
Institut Curie	2010	0	4149	4153	5458	3425
INHA		0	3495	9512	3976	3130
SCD Paris 1		0	1476	1738	1016	926
CNAM		0	846	842	428	196
SCD Angers		895 (P)	291	251	284	176
Ecole militaire Paris		0	191	28	906	792
SCD Paris 6		0	0	2637	0	0
BIU Cujas		2011	226	4203	4079	1239
ENS Ulm / CAPHES	4 (P)		11082	13500	6237	3233
SCD Maine	0		504	504	1120	283
SCD Tours	0		292	300	5	0
SCD Picardie	0		170	166	329	45
SCD Poitiers	2		45	40	171	153
Institut Catholique Paris	0		31	49	23	17
Ecole des Mines ParisTech	0		0	0	0	0
CIRDOC Béziers	2012	0	6018	6201	6258	2385
Centrale ECP		0	1232	1549	1141	77
Sciences Po Paris		985 (P)	985	0	1974	4
MMSH Aix		0	601	596	6347	5769
SCD Paris 7		0	0	22	0	0
Observatoire Paris		0	0	4713	0	0
SCD Rennes 1 et 2	2013	303	309	0	559	0
Lyon BDL (+ BIU LSH)		152	336	524	248	0
CRBC Brest	2014	2462 (P)	2462	0	12463	0
SCD Nice		130 (P)	130	0	424	2
SCD Franche-Comté		13 (C)	13	0	30	0
Etablissements non déployés au printemps 2014						
Clermont BCC	CGM	3192 (C)	3192	0	7623	0
Chantilly	CGM	2966 (C)	2966	0	9767	2
Archives Académie Fra.	CGM	815 (C)	815	0	806	0
Soc. Evreux	CGM	285 (C)	286	0	582	158
Ecole Polytechnique	CGM/Palme	178 (C+P)	133	0	120	64
SCD Perpignan	Palme	131 (P)	131	0	581	0
SCD Lille 3	CGM	104 (C)	104	0	372	0
Chalais	CGM	93 (C)	93	0	389	0

Etablissement	Année de déploiement	C publiés fin 2007	C publiés (mi mars 2014)	C créés dans l'outil (2008-2013)	Nb total de points d'accès	Nb points d'accès liés à IdRef
Ecole vétérinaire Alfort	CGM	71 (C)	71	0	174	0
SCD Caen	CGM	70 (C)	70	0	158	0
Musée Jacquemart André	CGM	61 (C)	61	0	168	0
Lyon Univ. Catholique	Palme	58 (P)	58	0	85	0
SCD Aix Marseille 3	CGM	53 (C)	53	0	75	0
SCD Metz	Palme	41 (P)	41	0	87	0
SCD Paris 3 (Bib. Gaston Baty)	Palme	33 (P)	33	0	45	0
SCD Valenciennes	Palme	16 (P)	16	0	64	1
SICD Grenoble 2	CGM	6 (C)	6	0	7	0
SCD Réunion	Palme	2 (P)	2	0	2	1
TOTAUX		158241	483097	443028	736092	242299

2. Evolution générale du nombre de <c> publiés dans Calames (2007-2014)



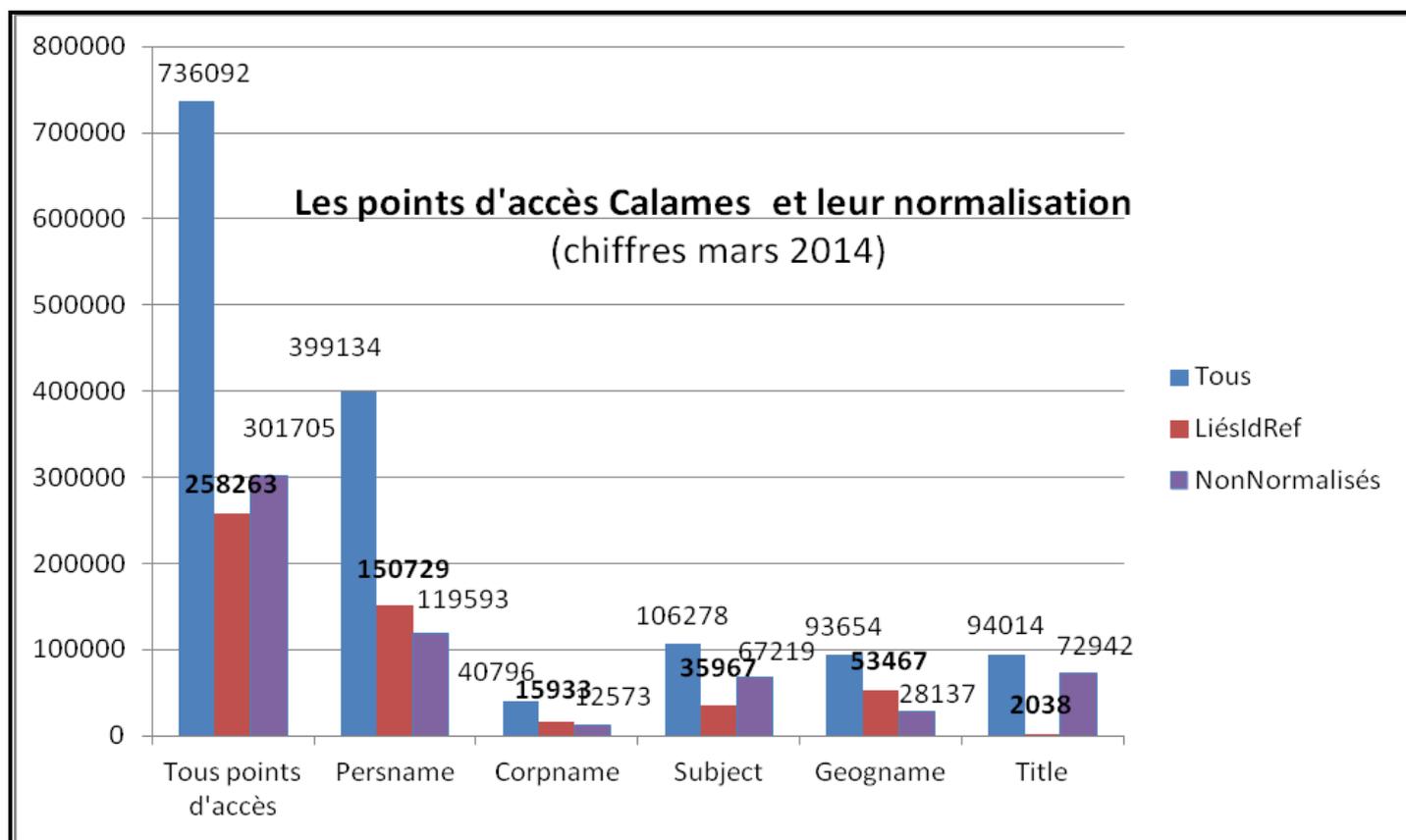
Voir également le blog Calames (statistiques des années 2012 et 2013) :

<http://calames.wordpress.com/2013/01/11/comptes-conte-dune-bonne-annee/>

<http://calames.wordpress.com/2014/01/09/statistiques-calames-2013/>

3. Statistiques relatives aux points d'accès dans Calames

[N.B. : chiffres portant sur les seules données *publiées* en mars 2014)

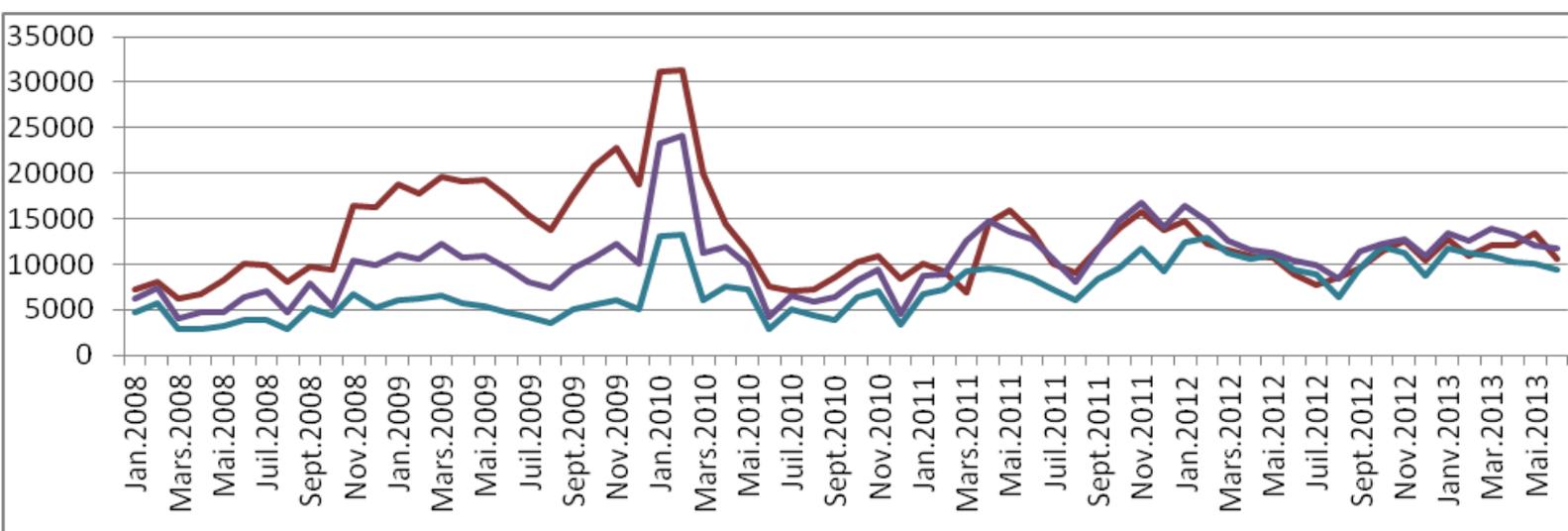


	Persname	Corpname	Famname	Subject	Geogname	Title	TOUS
Tous points d'accès	399134	40796	2216	106278	93654	94014	736092
Liés à IdRef	150729	15933	129	35967	53467	2038	258263
Non normalisés	119593	12573	1241	67219	28137	72942	301705

4. Statistiques de consultation (2008-2012)

Les données statistiques issues de Google Analytics et des logs enregistrés sur les serveurs de l'ABES sont à bien des égards divergentes, parfois même contradictoires. Une étude détaillée des cinq premières années d'existence de l'interface publique Calames a permis d'aboutir à cette conclusion : en se fiant aux seules données serveurs ABES, la quantité moyenne de visites mensuelles passe de 8000 à 12000 entre 2008 et 2012. La quantité des publics drainés par un page-ranking favorable de Google, massive autour de l'année 2009, a brutalement chuté au printemps 2010 et continue à s'éroder. Les publics avertis, qui utilisent Calames en tant que base et outil, croissent quant à eux d'une manière régulière. Le sur-référencement initial de Google a cédé la place à un sous-référencement certain, dans la mesure où Calames s'est vu retirer une part de son public pertinent (en même temps qu'une grande partie des visiteurs non intéressés ou passifs). Que Google n'entre plus dans l'acheminement vers les inventaires de Calames qu'à hauteur de moins d'un quart des visites (contre les trois quarts autour de 2009) doit-il être considéré comme satisfaisant ? En supposant que Calames a bien une vocation à être plus que l'antichambre des salles patrimoniales de son réseau, faut-il se fixer pour objectif d'être visible de meilleures façons, et d'insuffler davantage de sérénité dans les accès à Calames – eu égard notamment à une richesse de contenus en continuelle progression (nouveaux inventaires, liens hypertextes et accès à des versions numérisées...) ?

Recherches dans l'interface publique et diversité des accès à Calames



En rouge : Nombre d'adresses IP distinctes interrogeant Calames

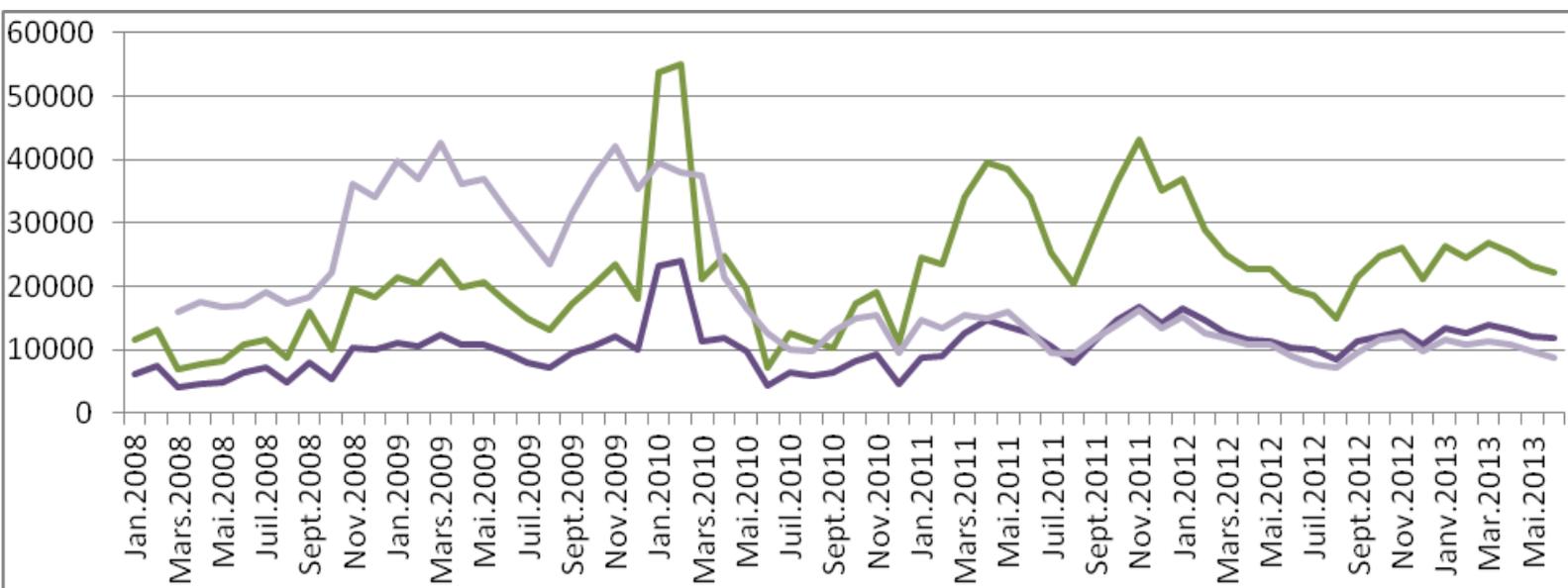
En violet : Nombre de recherches distinctes dans Calames

En bleu : Nombre de recherches distinctes sauf robots et ABES

Ces courbes, qui reposent sur les chiffres des seuls logs des serveurs ABES (et non sur les chiffres de Google Analytics), montrent le rôle « d'interface par délégation » que le moteur Google a plus particulièrement assuré de fin 2008 à début 2010 : le différentiel entre nombre d'adresses IP distinctes et nombre de recherches effectuées témoigne d'une large majorité d'accès simples aux notices Calames après recherches Google. Les fonctions de recherches de Calames, y compris la recherche plein texte (recherche simple), étant alors relativement sous-utilisées.

Les tendances d'évolutions du nombre d'IP (visiteurs mensuels) et du nombre de recherches effectuées sur les données de Calames se corroborent mutuellement.

Visites d'internautes et affichages détaillés de notices



En vert : nombre de notices vues, sauf robots et ABES [logs ABES].

En violet foncé : nombre visualisations de notices uniques, sauf robots et ABES [logs ABES].

En mauve clair : nombre de visites [d'après Google Analytics]

On observe une concordance remarquable à partir de mi-2010 entre pages vues d'après Google Analytics et nombre de visualisations distinctes de notices effectivement demandées par des utilisateurs humains, d'après les logs ABES. Cela porte à se méfier du très important succès public que les statistiques Google Analytics semblent indiquer en 2008-2010. Les modifications dans les modes de comptage et dans les algorithmes de recherche de Google expliquent en partie de telles variations.

5. Statistiques Calames Plus Images (chiffres mars 2014)

		Musée Curie	Bib. Litt. J. Doucet	Bib. Mazarine	TOTAL
[Stats. dépôt]	Nombre d'images déposées	791	1506	7094	9391
[Stats. dépôt]	Poids déposé (Go)	0,468	3,45	42,3	46,3
[Stats. dépôt]	Nombre de fichiers (dont tuiles)	35165	356355	2141512	2533033
[Stats. dépôt]	Nombre de <c> indexés (index "dao")	791	185	31	1007
[Stats. visionneuse]	Images vues (cumul 06/2012-02/2014)	4503	6195	10602	21357
[Stats. visionneuse]	Images vues (moyenne mensuelle)	225	310	530	1068

N.B. : le nombre total d'adresses IP distinctes ayant consulté la visionneuse Calames Images entre juin 2012 et février 2014 s'élève à 1438. Chaque mois, on compte en moyenne 72 adresses IP distinctes accédant aux images déposées.