

Point de vue d'historienne de l'art sur les usages actuels et prospectifs de Calames

Je vais rapidement me présenter : je suis docteure en histoire de l'art, spécialiste de l'art français du XIX^e siècle et de l'histoire institutionnelle et sociale du Salon. Dans le cadre de mes recherches, je consulte davantage les Archives nationales et leurs outils, puisque ce sont là que sont conservées la plupart des archives avec lesquelles je travaille. Néanmoins, j'utilise ponctuellement Calames pour retrouver des archives privées et surtout, cela m'intéresse de plus en plus dans une optique de recherche historiographique.

Pour la commande du jour, je propose de naviguer à travers les différents usages possibles de Calames, en découpant chaque tâche et action afin de voir ce qui pourrait être fait ou amélioré. Les premiers usages sont les plus anciens et traditionnels : trouver des archives - probablement les utilisations les plus fréquentes de vos utilisateurs et utilisatrices, ce qui explique certainement la grande majorité des choix d'interface effectués dans la version actuelle. Je propose également une approche prospective en évoquant de potentiels futurs usages qui doivent être considérés dès maintenant pour les rendre possibles et ne pas vous couper l'herbe sous le pied.

1. Identifier les archives

L'usage le plus courant chez les chercheurs et chercheuses de Calames est d'y trouver leurs archives, ce qui comprend en réalité deux étapes : les identifier puis les localiser.

Pour identifier et repérer les archives, il faut déjà interroger la base de données et donc lui soumettre une requête. Calames permet de requêter avec une recherche simple et une recherche plus avancée, ce qui offre la possibilité de croiser les critères de recherche. Un reproche néanmoins peut être fait aux deux types de recherche : l'algorithme n'est pas explicite et n'indique pas sur quel champ se fait la recherche « générale ». Concrètement, quelles sont les métadonnées qui vont être requêtées quand on cherche par mots-clés ? Seulement le champ de description ou aussi les intitulés ? (ce qui peut alors générer des « fausses réponses »). Il serait préférable d'avoir un formulaire de recherche très explicite, à l'instar du formulaire de recherche avancée de la Bibliothèque nationale de France, avec des connecteurs "et", "ou", "sauf" visibles et des champs d'application rendus explicites qui permettent une requête sur mesure.

En outre, il ne faut pas oublier certains champs dans le formulaire, notamment la cote : les chercheurs et chercheuses qui ont déjà repéré une archive ailleurs (dans les références de monographies, dans des guides de recherches ou dans d'autres bases de données, par exemple GAEL de l'Institut national d'histoire de l'art) peuvent vouloir accéder directement à la notice correspondante en utilisant la cote.

Une fois la requête lancée, l'affichage des résultats s'avère globalement fonctionnel. On peut néanmoins regretter l'absence de certaines informations qui pourraient considérablement faciliter l'identification des archives pertinentes pour le chercheur et la chercheuse. Si les termes et les opérateurs de la recherche sont rappelés – ce qui est toujours utile –, ce qui serait encore mieux serait de voir apparaître le contexte des résultats trouvés, *a fortiori* quand la recherche par mots-clés s'applique à l'ensemble des champs décrivant le document. C'est par exemple ce que propose la Salle des inventaires virtuels des Archives nationales ou France Archives, ce qui permet en un rapide coup d'œil de voir l'intérêt de l'archive suggérée – ce qui, au demeurant, ne dispense pas la plupart des chercheurs et chercheuses de quand même ouvrir la notice. Sans cette mise en contexte, il peut parfois être compliqué de comprendre les résultats obtenus, d'autant que le caractère très voire trop général du moteur de recherche, et son absence d'explicité, fait parfois tomber sur des résultats anecdotiques.

Le filtrage des résultats pourrait aussi être largement amélioré, à deux niveaux. D'une part, en corrigeant et précisant les indexations proposées : certains termes pourraient être regroupés lorsqu'ils ont des significations très proches, d'autant qu'il n'est pas possible d'en sélectionner plusieurs. En outre, certains champs pourraient être rajoutés comme critères de filtrage : en particulier, les dates de production – particulièrement discriminantes pour les historiens ou historiennes – ou les producteurs, quitte à ce que ceux-ci ne soient pas identifiés individuellement mais par types (archives privées individuelles ou de personnes morales, officielles ou non). Il faut en outre, dès qu'il y a possibilité de filtre, permettre de sélectionner plusieurs valeurs donc, mais aussi de réinitialiser ou du moins d'enlever des critères de filtre, sans avoir refaire la recherche en entier.

L'obtention de résultats à différents niveaux d'archives, du fonds au document en passant par les dossiers et sous-dossiers, est extrêmement utile : elle assure aux chercheurs et chercheuses, qui ne sont pas nécessairement lettrés en archivistique, de pouvoir se repérer et de s'assurer de ne rater aucun niveau d'organisation potentiellement intéressant. On peut toutefois regretter que cette précision, évidemment précieuse, empêche parfois de saisir les ensembles qui figurent dans un même lieu ou qui constituent en réalité un fonds en soi : à ce titre, la solution mise en place par la Salle des inventaires virtuels, qui donne le choix à l'utilisateur de

regrouper ou non les documents, dossiers et fonds appartenant à un même inventaire, pourrait être reprise. Dans les notices elle-même, le fait que l'arborescence archivistique réapparaisse est extrêmement précieuse : elle permet d'une part au chercheur de comprendre l'organisation interne du fonds ; de pouvoir potentiellement remonter et naviguer dans d'autres dossiers et sous-dossiers susceptibles de l'intéresser et de localiser la place occupée par le document dans l'archive en général – notamment en vue de la citer le plus précisément possible. Le panneau latéral de navigation, qui permet de circuler dans les notices précédentes et suivantes, et de revenir à la liste des résultats, est satisfaisant mais pourrait peut-être encore plus visible. On m'a toutefois part d'une gêne récurrente : celle de ne pas pouvoir ouvrir, à partir de la liste des résultats, chaque notice dans un onglet – une pratique extrêmement fréquente – ou de ne pas pouvoir revenir aux résultats par le bouton « Page précédente » du navigateur – qui est le réflexe des utilisateurs. Laisser cette possibilité aux utilisateurs, c'est leur assurer une navigation sans trop de frustration et donc une meilleure mise en valeur de vos fonds.

Outre l'affichage des résultats, il convient aussi de rendre possible pour les chercheurs et chercheuses de conserver ces résultats. Si à l'heure actuelle la recherche comme les notices disposent d'URL unique, qui peuvent donc être gardés soit en liens hypertextes soit en signets sur le navigateur, il pourrait être utile d'envisager la possibilité pour les utilisateurs et utilisatrices de se constituer des « panier de cotes » dans lequel ils puissent revenir aisément.

2. Localiser les archives

Une fois les fonds et documents identifiés comme pertinents pour nos recherches, il nous faut donc localiser les archives dans l'optique d'aller les consulter. Actuellement, le portail Calames ne présente les différentes institutions cataloguées que par ordre alphabétique de la ville : c'est un mode de navigation pratique pour qui sait *déjà* où il compte aller, mais assez peu pratique pour une personne voulant rentabiliser son temps et son énergie. Les chercheurs et chercheuses prévoient en effet leur séjour de recherche en fonction de la pertinence des archives à consulter évidemment, mais aussi d'un point de vue pratique : quitte à partir dans une autre région, autant regrouper en un même séjour toutes les institutions à proximité. À ce titre, si cela peut paraître un peu ésotérique, il n'est en réalité pas complètement absurde de proposer aux chercheurs et chercheuses une cartographie des institutions présentes dans Calames, comme le propose pour les objets du patrimoine la plateforme POP. Cela peut permettre aussi de donner une idée de la concentration d'archives pertinentes dans un même lieu ou une même région et alors de faire émerger de potentielles nouvelles questions et découvertes : pourquoi ? est-ce le signe d'un intérêt d'un acteur local à mieux documenter ?

Il faudrait aussi penser aussi à faciliter la consultation en offrant de meilleures informations sur les institutions. Actuellement, les notices par document ou fonds ne rendent pas très explicite le lieu où se trouve l'institution : ici par exemple, il n'est indiqué que « Bibliothèque littéraire Jacques Doucet » et il faut que je clique pour apprendre qu'elle se situe à Paris ... On pourrait envisager de faire figurer la ville et de permettre en outre, en cliquant sur l'institution, de faire apparaître une nouvelle fenêtre ou un nouvel onglet (mais qui ne s'ouvre *pas* sur la même page pour ne pas faire perdre la notice consultée) qui précise les informations pratiques, et notamment les horaires d'ouverture et les conditions de consultation. Encore une fois, cela permettrait à Calames de devenir véritablement l'outil central pour les chercheurs et chercheuses et qu'ils n'aient alors quasiment plus besoin de retourner, individuellement, sur le site de chaque institution pour obtenir ces informations.

3. Extraire les métadonnées

Passons à un usage plus prospectif : il convient en effet d'envisager d'autres utilisations de Calames que la simple identification et localisation d'archives, mais par exemple un moissonnage plus ample des données contenues pour faire des traitements quantitatifs et statistiques. À l'heure actuelle, Calames ne permet aucune extraction des métadonnées – du moins, pas de façon suffisamment claire et simple pour que je l'ai vu. La liste complète des résultats ne peut être téléchargée ou exportée sous aucun format – PDF, CSV ou autre. Idem pour les notices individuelles : pas de possibilité de les extraire. Or, je pense qu'il devient désormais nécessaire de proposer *a minima* différents types d'extraction, selon les potentiels usages qu'en font les chercheurs et chercheuses.

À l'instar de la salle des inventaires virtuels des Archives nationales, il serait utile – pour ne pas dire indispensable – de permettre de télécharger l'inventaire des fonds en PDF : la pratique peut vous paraître archaïque ou obsolète, mais pour avoir mener mon enquête auprès de collègues, nombreux sont encore les chercheurs et chercheuses à aimer posséder une version PDF des inventaires les plus susceptibles de les intéresser et/ou qui comprennent un grand nombre d'archives pertinentes pour leur recherche, afin de faire des recherches plein texte notamment. C'est aussi un document précieux à conserver sur un temps long – moyen disons – afin de garder traces des fonds consultés mais aussi parce que la recherche fonctionne de façon tentaculaire et qu'il est souvent précieux et bénéfique de ne pas entrer dans les fonds que par une recherche précise, mais en se laissant voguer et consulter les cotes voisines. Personnellement, j'aime aussi posséder les inventaires en PDF des fonds que je consulte le plus souvent, que je stocke dans ma bibliothèque Zotero, pour pouvoir y revenir au fur et à mesure

des approches qu'adopte ma recherche. C'est donc là un premier format d'extraction des données contenues dans Calames – soit au niveau des inventaires seuls, soit plus largement de la liste des résultats – qu'il me paraît nécessaire d'autoriser.

Dans une perspective davantage quantitative, il pourrait aussi être utile de proposer une extraction des données au format CSV, comme le propose le portail France Archives. Ici, la possibilité est donnée de télécharger la notice elle-même : c'est déjà une chose utile, puisque le chercheur ou la chercheuse pourra alors collecter dans un même tableur les métadonnées des notices réunies. Le catalogue de la BNF permet quant à lui de collecter en lot les notices, soit en effectuant une sélection sur les résultats proposés soit en moissonnant l'ensemble des résultats : là encore, il est possible de télécharger la liste obtenue au format PDF mais aussi de l'exporter dans un tableur, c'est-à-dire au format CSV. Cette possibilité permettrait alors aux chercheurs et chercheuses d'exploiter les métadonnées de plusieurs notices pour les soumettre à des traitements statistiques par exemple et pour leur poser des questions peut-être davantage historiographiques ; en gros, ne plus seulement employer les archives comme des sources mais aussi comme des objets d'études en soi, qui peuvent raconter des choses intéressantes sur la manière dont certains phénomènes ou acteurs laissent des traces (ou non), passent la sélection de la postérité (ou non), sont suffisamment représentés pour être étudiés (ou non) ... En permettant d'extraire notice par notice ou en lot (c'est encore le mieux) des tableaux contenant les métadonnées des archives et des fonds, il s'agirait ainsi de permettre une valorisation différente, mais tout aussi pertinente, des fonds conservés par le réseau Calames.

Dans une même optique, on peut aller encore plus loin en proposant, en sus, une extraction à des formats lisibles par les machines directement, comme le XML, le RDF ou le format JSON (ce sont des options proposées par les Archives nationales ou la BNF par exemple). Il n'y pas nécessairement besoin de proposer des interfaces très sophistiquées : une simple interface API, qui propose chaque notice au format JSON, peut se révéler largement suffisante sans alourdir le portail de consultation pour les chercheurs et chercheuses. Ces formats, qui ne sont pas nécessairement utilisés par la majorité des chercheurs et chercheuses, peuvent néanmoins être moissonnés par les machines et des algorithmes et pourraient ainsi permettre aux métadonnées qui décrivent les archives conservées par le réseau Calames d'être exploités par d'autres programmes de recherches, dans une perspective d'interopérabilité.

4. Exploiter les métadonnées

Si permettre l'extraction des métadonnées peut constituer en soi une amélioration forte pour le portail Calames, encore faut-il s'en assurer une exploitation optimale. Il s'agit à ce titre

de constituer un modèle de description des données le plus cohérent et homogène possible de chaque document (fonds et pièces d'archives) faisant l'objet d'une notice, ce qui n'est pas tout à fait le cas aujourd'hui. En guise de preuve, prenons le champ « Description physique » : dans les quatre notices que je prends pour exemple ici, on remarque que ce champs peut être rempli par le nombre de feuillets conservés ; par le type de documents contenus ; par des dimensions ; enfin par des mètres linéaires, déjà plus orthodoxes en milieu archivistique. Pour un chercheur ou une chercheuse intéressé par la présence, dans les archives, d'une personnalité quelconque, cette hétérogénéité des modes de description est problématique et empêchera tout traitement statistique et comparatif par exemple. Si je comprends bien la difficulté, voire l'impossibilité, de décrire de façon identique la diversité et la pléthore de documents conservés par le réseau Calames, il pourrait néanmoins être utile de s'accorder sur des vocabulaires contrôlés ou des modèles précis de description afin que les métadonnées soit le plus facilement et utilement exploitables possibles.

Plus encore, il pourrait être utile d'adopter une sémantisation des métadonnées contenues, à l'instar de ce que propose France Archives. On remarque en effet que les individus concernés par les archives, qu'ils en soient les producteurs et qu'ils soient mentionnés, font l'objet d'une notice spécifique, ce qui a plusieurs avantages. Premièrement, cette notice individuelle permet de regrouper en un même endroit toutes les archives concernant cet individu. Deuxièmement, elle permet d'obtenir des informations de base sur cet individu, comme les dates de naissance et de mort ainsi que la nationalité, ce qui peut permettre de discriminer et de faire le tri dans le cas d'homonymie. Troisièmement, elle réfère à des notice d'autorités et à des sources (wikidata et data.bnf le plus souvent) ce qui facilite plus encore l'interopérabilité de ces différentes données. Une telle sémantisation pourrait alors faciliter des études sur, par exemple (je dis n'importe quoi) les femmes dans les archives bretonnes, en moissonnant l'ensemble des archives conservées en Bretagne, en alignant les personnalités concernées à leur référentiel et en déterminant ensuite le sexe ratio. L'effort de sémantisation peut aussi concerner les lieux ou encore les thèmes traités ; bref tout ce qui peut assurer aux chercheurs et chercheuses de travailler sur des métadonnées de qualité, c'est-à-dire indexées, référencées, accessibles et utilisables sans trop de manipulation supplémentaire (de correction ou d'homogénéisation notamment).