

Archives sur Internet : quels rôles pour l'archiviste ?

Sarah Cadorel

Citer ce document / Cite this document :

Cadorel Sarah. Archives sur Internet : quels rôles pour l'archiviste ?. In: La Gazette des archives, n°239, 2015-3. Chemins de traverses : ces métiers au service des archives. Regard d'une ethnologue. pp. 141-149;

doi : <https://doi.org/10.3406/gazar.2015.5339>

https://www.persee.fr/doc/gazar_0016-5522_2015_num_239_3_5339

Fichier pdf généré le 12/05/2018

Archives sur Internet : quels rôles pour l'archiviste ?

Sarah CADOREL

Quel regard portons-nous sur les archives et sur le métier d'archiviste lorsque notre activité est centrée sur la gestion d'un site Internet de mise à disposition ? La mise en ligne d'archives aujourd'hui ne touche pas seulement les centres d'archives historiques. Elle fait l'objet de nombreux projets dédiés à la recherche scientifique, qui s'intéressent à la valeur de réutilisation des archives. Ces archives, nativement numériques ou numérisées – souvent appelées données – soulèvent un certain nombre de questionnements quant à leur communication sur Internet. La gestion d'un site Web de mise en ligne d'archives comprend sa conception, son enrichissement, sa maintenance et son évolution. Être archiviste au sein d'une organisation dont l'activité est centrée autour d'un site de mise à disposition d'archives¹ influence la manière d'exercer son métier et de se situer dans son appartenance professionnelle. En conciliant contraintes technologiques, exigences scientifiques et spécificités propres aux archives, l'archiviste doit faire preuve de souplesse pour rendre les archives accessibles, lisibles et intelligibles dans tout ce qu'elles ont de particulier. En effet, si l'environnement Web impose un certain nombre de règles à suivre, en termes de construction, de procédure, d'interlocuteurs et de terminologie, il en est de même pour les archives qui, par leur nature organique, leur structure arborescente et leur unicité se plient difficilement à la logique binaire de l'informatique. Dans quelle mesure la mise en ligne d'archives affecte-t-elle la pratique archivistique ? Si nous sommes amenés à déroger à un certain nombre

¹ L'auteur est actuellement archiviste au sein d'un projet de mise à disposition d'archives d'enquêtes qualitatives en sciences sociales nommé beQuali. Ce projet constitue l'un des trois volets de l'équipement d'excellence DIME-SHS (données, infrastructures et méthodes en sciences humaines et sociales) porté par le Centre de données socio-politiques (UMS 828 Sciences Po/CNRS).

de principes, comme la logique de producteur ou la description arborescente, nous devons acquérir des compétences indispensables pour mener à bien les projets. Comment alors envisager le métier d'archiviste et l'avenir professionnel¹ ? Notre propos sera centré sur les archives de la recherche, objet des infrastructures de mise à disposition récentes, et ne donnera qu'une vision partielle du champ des ressources archivistiques disponibles sur la toile.

Archives à disposition

Plus qu'un outil au service de la conservation préventive ou qu'un instrument de communication au service de la valorisation des activités d'un service, les sites Web de mise en ligne d'archives sont devenus de véritables infrastructures d'aide à la recherche. Le traitement des archives n'intéresse plus seulement les historiens mais des chercheurs d'autres disciplines, telles que la sociologie, la géographie ou encore la linguistique, qui souhaitent disposer de matériaux scientifiques dans des délais de plus en plus courts.

Le carcan du numérique

Quel que soit son contenu, un site Internet est construit selon un certain nombre de règles et de normes concernant aussi bien son architecture que son fonctionnement. D'une part, les contraintes du système informatique dans lequel il est hébergé imposent une adaptabilité réciproque entre le contenu et l'architecture matérielle et logicielle. Les limites de poids sont, par exemple, un élément à prendre en compte, en particulier lorsque celui de fichiers issus de la numérisation au format tiff² ou jpeg 2000³ peut atteindre plusieurs dizaines de mégaoctets. Afin ne pas saturer le réseau, le format des fichiers doit être optimisé pour leur chargement, quitte à en diminuer la qualité. Les formats de fichiers ou la structure de leurs identifiants doivent également être détectables et intelligibles par un système technologique choisi répondant à un besoin guidé par la nature

¹ Sur ce sujet la journée d'études « Les archivistes dans quinze ans » organisée par l'université de Louvain-la-Neuve (Belgique), les 23 et 24 avril 2015, atteste de l'évolution en cours de la profession d'archiviste.

² <http://partners.adobe.com/public/developer/tiff/index.html#spec>

³ http://www.bnf.fr/fr/professionnels/numerisation_boite_outils/a.numerisation_referentiels_bnf.html

des documents traités et l'exploitation qui veut en être faite. D'autre part, les usages et les normes en termes de navigation, d'ergonomie et de graphisme conditionnent les fonctionnalités développées dans une application Web. La littérature professionnelle informatique nous recommande par exemple d'éviter l'utilisation d'effets *pop up* jugés peu ergonomiques pour l'ouverture de nouvelles fenêtres. Aussi, la règle des « trois clics » impose au concepteur de site Web de ne pas faire parcourir plus de trois niveaux de menus à l'utilisateur pour accéder à l'information.

Si ces règles génériques s'appliquent à toute application sur Internet, il est des spécificités qui incombent aux sites Web documentaires. En effet, le cœur de telles applications est généralement composé d'un moteur de recherche qui permet d'effectuer des requêtes selon des critères adaptés à la nature du *corpus* hébergé. La pertinence des résultats issus des moteurs de recherche reste le point d'achoppement des plateformes de mise à disposition car elle garantit le succès du dispositif. Tel est d'ailleurs l'objectif poursuivi par les évolutions du Web sémantique qui entend proposer aux utilisateurs des résultats toujours plus pertinents en adaptant les moteurs de recherche au plus près des objets numériques contenus dans les sites Internet¹.

En outre, la grande spécificité des archives réside dans leur caractère organique et dans leur structure arborescente. Mettre en ligne un *corpus* archivistique est plus complexe qu'une base de données « à plat » où tous les objets se situent au même niveau de classement. Ici, le tout n'est pas la somme des parties et il s'agit de concilier un carcan numérique, borné par des choix technologiques, et un contenu à l'organisation à la fois composite et sédimentaire que sont les fonds d'archives.

L'archivistique malmenée

Si l'on excepte les difficultés liées à la définition du processus de numérisation ou à la maîtrise des technologies de l'image numérique, mettre en ligne des collections ou des séries de documents regroupées sur des critères de forme ou de nature ne semble pas poser de difficulté particulière. C'est ainsi que les bases de données ont fleuri et possèdent désormais en leur sein une masse documentaire impressionnante. Tel est le cas du répertoire des manuscrits reproduits et recensés Medium² de l'Institut de recherche et d'histoire des textes

¹ POUYLLAU (Stéphane), « Les moteurs de recherche profitent aussi de la sémantique », *Documentaliste - Sciences de l'Information*, ADBS, n° 48 (4), 2011, p. 36-37. <halshs-00741328>

² <http://medium.irht.cnrs.fr/>

(IRHT) ou encore les registres d'état civil¹ mis en ligne par les Archives nationales, départementales et communales, qui constituent une volumétrie considérable de documents, aux caractéristiques diplomatiques homogènes et répartis sur une large étendue territoriale et temporelle.

À l'inverse, rendre compte de l'intégrité intellectuelle et matérielle de *corpus* beaucoup plus réduits mais à la structure complexe, tels que les archives de la recherche, dans un environnement virtuel, est ardu. En effet, leur production extrêmement personnelle, car peu régie par des procédures administratives, fait qu'une large part de leur valeur scientifique et historique réside dans la manière dont ils ont été constitués et organisés. Cela en dit long sur la méthode de travail des producteurs, sur les événements qui ont pu influencer le processus de recherche. Le défi est donc de concilier un traitement archivistique en silo suivant une logique de producteur et le fonctionnement résolument transversal du Web basé sur le partage et les échanges². Pour assurer l'interopérabilité des applications, leur moissonnage et leur visibilité, les paquets d'informations mis en ligne doivent être structurés et normés. Les documents doivent être convertis dans des formats adaptés à Internet, leur assemblage doit s'insérer dans l'architecture informatique. Les informations descriptives qui leur sont attachées et qui permettront de les identifier, de les décrire et de les gérer doivent être traduisibles en différents langages. La pratique archivistique s'en trouve alors ébranlée par la réduction de la souplesse des catégories et des niveaux de classement, ou par l'imposition de vocabulaires contrôlés allant parfois à l'encontre de la terminologie employée par les producteurs. Pour pallier cela, nous essayons d'enrichir la description par l'utilisation de plusieurs standards de métadonnées qui s'additionnent et se complètent. L'objectif est de faciliter les échanges avec d'autres plateformes de mise à disposition et de répondre aux besoins des utilisateurs dont les exigences scientifiques diffèrent de celles des lecteurs traditionnels des services d'archives. Même si ISAD(G)³ et son implémentation EAD⁴ sont irremplaçables pour décrire l'arborescence des documents, elles ne peuvent se substituer à des normes métier qui permettent de documenter les aspects très spécifiques mais primordiaux pour les chercheurs d'une discipline donnée. Ainsi, les sites Web de *corpus* linguistique, comme

¹ <http://www.archivesdefrance.culture.gouv.fr/ressources/en-ligne/etat-civil/>

² ANGJELI (Anila), « La normalisation en bibliothèque à l'heure du web sémantique », *Documentaliste - Sciences de l'Information*, ADBS, n° 48 (4), 2011. Disponible en ligne : http://www.cairn.info/article.php?ID_ARTICLE=DOCSI_484_0042&DocId=45005&Index=%02Fcairn2Idx%02Fcairn&T.<hal-00875829>

³ Norme internationale de description archivistique : [http://www.icacds.org.uk/fr/ISAD\(G\).pdf](http://www.icacds.org.uk/fr/ISAD(G).pdf)

⁴ *Encoded Archival Description* (<http://www.loc.gov/ead/>)

Ortolang¹, utilisent le standard OLAC² pour documenter les extraits sonores, tandis que les sites s'adressant aux chercheurs en sciences sociales, comme beQuali³, se servent du standard DDI⁴ pour documenter la méthodologie de l'enquête de terrain ou encore les géographes se conforment à la directive INSPIRE⁵ pour la mise en ligne des données spatiales.

En conduisant donc l'archiviste à modifier ses pratiques, la diffusion numérique opère un glissement de la logique de producteur vers une logique d'utilisateur.

Un dispositif orienté vers le « service »

Sur un site Web, le public est qualifié d'utilisateur tandis qu'on choisira plutôt le terme de lecteurs pour désigner les individus qui fréquentent les salles de consultation des services d'archives. Cela n'est pas anodin car, par le biais du téléchargement, un utilisateur peut s'approprier les documents mis en ligne, en les classant et en les déclassant à sa guise, en les renommant ou encore en les scindant par le jeu du copier-coller⁶. Nous parlons de mise à disposition et non plus seulement de communication, car le comportement des usagers en ligne est résolument différent. En effet, le terme d'utilisateur est calqué sur les portails de diffusion de données, telles que celles qui sont diffusées par réseau Quetelet⁷ ou par le Centre de données astronomiques de Strasbourg⁸, dont la production est coûteuse et rend indispensables le partage et la mise à disposition à la communauté scientifique à des fins de réutilisation. Si ladite communauté s'intéresse non plus aux seules informations mais aux archives dans leur globalité, elle émet les mêmes attentes en termes de service rendu de la part des diffuseurs. Le service rendu au public s'évalue ainsi par la proximité temporelle et spatiale d'accès aux documents avec un délai de consultation souhaité proche de l'immédiateté depuis son ordinateur. Paradoxalement, le public n'a jamais été aussi distant de l'archiviste. En effet le rapprochement à la fois intellectuel, par le croisement des informations, et géographique, par les déplacements qu'il évite

¹ <https://www.ortolang.fr/#/market/news>

² <http://www.language-archives.org/>

³ www.bequali.fr

⁴ *Data Documentation Initiative* : <http://www.ddialliance.org/>

⁵ <http://inspire.ec.europa.eu/index.cfm>

⁶ BOTH (Anne) et CADOREL (Sarah), « En quête d'enquêtes. La collecte auprès des chercheurs en SHS : le cas de beQuali. », *Actes du colloque de l'ICA*, Paris 8-10 juillet 2014 (à paraître).

⁷ Le réseau Quetelet est l'une des composantes de la très grande infrastructure de recherche du CNRS Progedo (<http://www.reseau-quetelet.cnrs.fr/spip/?lang=fr>)

⁸ <http://cds.u-strasbg.fr/>

aux chercheurs, renforce l'accessibilité des archives, il jette un voile sur le traitement archivistique. Dès lors, il est nécessaire de prévoir des garde-fous afin de prévenir les risques d'utilisations pernicieuses. Cela peut prendre la forme de métadonnées descriptives, notamment métier qui, par leur richesse, peuvent limiter les risques de décontextualisation ou de faux-sens lors de la réutilisation des informations contenues dans les documents. De plus, les métadonnées techniques permettent de marquer les documents d'une sorte d'empreinte indélébile assurant leur traçabilité. L'anonymisation s'avère être également un recours pour limiter les risques liés à la diffusion de données personnelles et sensibles pouvant porter préjudice aux personnes et sur lesquelles la Commission nationale informatique et libertés (CNIL) reste intransigente.

Le métier d'archiviste

La pratique archivistique se caractérise par plusieurs éléments : un objet, une déontologie, une méthode et une finalité. Placé au cœur d'une équipe aux métiers multiples dont les activités sont centrées sur l'aide à la recherche, l'archiviste est amené à diversifier ses activités.

La réalité du travail au quotidien

Au quotidien, les activités d'un archiviste dans une unité au service de la recherche recouvrent une réalité singulière. Positionné sur un projet, il s'agit de composer avec des profils, des compétences et des logiques professionnelles extrêmement différents. En effet, le dialogue entre développeurs, ingénieurs système, chercheurs, archivistes et documentalistes est loin d'être évident. Chacun possède un vocabulaire spécifique, même pour désigner une chose identique, et le coordinateur du projet doit développer des talents de traducteur indispensables. Les exemples de termes polysémiques ne manquent pas, à commencer par celui d'archivage. Aussi est-il prioritaire que les interlocuteurs puissent s'entendre, non seulement sur le sens des mots eux-mêmes mais aussi sur la réalité qu'ils désignent. Ainsi, la numérisation, terme compréhensible de tous, inclut une série d'opérations que beaucoup ignorent. Pour l'archiviste, comme pour chacun des acteurs précédemment cités, conduire un projet de numérisation d'archives nécessite une vraie montée en compétences et le passage de l'état original à celui de document numérisé suppose une parfaite maîtrise du

processus. Que la numérisation soit effectuée en interne par l'équipe du projet ou par un prestataire extérieur, la première étape consiste à rédiger un cahier des charges spécifiant tous les aspects techniques et organisationnels de l'opération. Cette étape clé conditionne la bonne conduite des opérations dont la qualité de réalisation sera évaluée lors d'une étape de contrôle final. Toutes ces tâches supposent des connaissances approfondies en technologies d'images numériques car il faut être capable de définir puis de contrôler les fichiers numérisés. Elles demandent également une maîtrise des normes de métadonnées car définir un schéma XML et contrôler un fichier au format METS¹, par exemple, n'est pas évident. Enfin, elles requièrent une méthode à la fois en termes de gestion de projet – surtout s'il s'agit de gérer une prestation à caractère contractuel – et de contrôle qualité. En outre, la conception, la maintenance et l'évolution de l'infrastructure elle-même mobilisent un travail de webmestre pour la définition des fonctionnalités du site et des processus nécessaires à leur mise en œuvre.

Rédiger des cahiers des charges, définir les *workflows*, solliciter les intervenants sur le projet sont autant de tâches qui incombent à l'archiviste œuvrant pour un projet de site Web au détriment de la rédaction d'instruments de recherche. Focalisée davantage sur la rédaction de spécifications fonctionnelles ou techniques que sur celle d'instruments de recherche, l'activité de conception et d'alimentation d'une plateforme de mise en ligne d'archives demande un niveau de compétences technologiques et un sens de la communication nouveaux. Pour autant, la finalité du travail de l'archiviste s'en trouve-t-elle transformée ?

Patrimonial, scientifique et technique

Les sites de mise à disposition d'archives ne peuvent être universels et sont forcément orientés vers un type d'usage particulier, surtout si la démarche du projet est guidée par une logique scientifique aux enjeux théoriques ou méthodologiques et non par la seule sauvegarde du patrimoine. L'inclusion d'utilisateurs finaux dès la conception des applications Web permet de mieux anticiper les attentes du public. Mais cela oriente également le choix technologique et les développements pour construire l'outil de diffusion des documents. Aussi pouvons-nous prendre comme exemple les *consortiums* labellisés Huma-Num² dont chaque équipement est développé pour une communauté scientifique particulière. L'archiviste exerçant au sein de ces projets est alors immergé dans une démarche scientifique et fait l'objet d'une

¹ *Metadata Encoding and Transmission Standard* : <http://www.loc.gov/standards/mets/>

² <http://www.huma-num.fr/service/consortium>

acculturation disciplinaire indispensable à l'exercice de son activité de service public. En effet, le classement des documents requiert une bonne compréhension tant de la forme que du contenu des matériaux, et l'utilisation de normes de métadonnées « métier » pour la description des archives suppose une relative maîtrise des enjeux scientifiques de la discipline concernée. De plus, rendre accessibles et réutilisables les documents implique d'assurer leur conservation au moins à moyen terme et d'assurer la pérennité de l'infrastructure qui les héberge. Pour cela, des compétences en archivage pérenne sont nécessaires. Dès lors la logique scientifique de réutilisation des archives rejoint la logique patrimoniale sans laquelle les documents ne seraient disponibles que de manière éphémère, et la clé de voûte du dispositif reste la maîtrise des technologies numériques et de leur mise en œuvre. Ainsi, l'archiviste remplit-il sa mission de service public par l'expertise qu'il apporte pour la sauvegarde du patrimoine scientifique qui en étant pérenne pourra être accessible à tous, en respect les délais de communicabilité et les recommandations de la CNIL. En effet, sa connaissance du réseau des archives en France est la garantie de la bonne conservation des originaux qui auront servi pour la mise en ligne, et ses compétences en archivage électronique pourront assurer la pérennité et la traçabilité du site Web et de son contenu.

Perspectives d'évolutions

Les archivistes de l'enseignement et de la recherche sont rattachés à deux ministères, celui de la Culture et de la Communication, et celui de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche. De fait, on observe une sorte de décalage dans les intitulés des postes entre les archivistes de la première administration appelés « conservateurs du patrimoine », « assistants de conservation » ou encore « chargé d'études documentaires » et ceux de la seconde qui sont « ingénieurs d'études » ou de « recherche ». S'il s'agit du même métier pour lequel les membres ont reçu la même formation initiale, par la suite l'exercice de la profession au sein d'un organisme et d'une équipe œuvrant pour la mise à disposition d'archives sur Internet mobilise une multiplicité de compétences : coordination et gestion de projet ; expertise technique pour les formats de fichiers le classement et la description ; expertise juridique dans un domaine complexe où plusieurs régimes de droits s'intercalent ; expertise scientifique sur les archives. L'aspect résolument informatique du travail de mise en ligne nécessite une montée en compétences dans le domaine de l'ingénierie documentaire. Aussi sommes-nous témoins d'une évolution du métier vers une technicisation qui semble s'éloigner de la science auxiliaire de l'histoire qu'est

l'archivistique, telle qu'elle est transmise lors des formations initiales. Par conséquent, il existe un véritable enjeu pour les archivistes de se situer au croisement de plusieurs domaines scientifiques et opérationnels. En témoigne la révision des référentiels métiers en cours au Centre national de la recherche scientifique (CNRS) ou au ministère de l'Éducation nationale, de l'Enseignement supérieur et de la Recherche, car l'emploi type ne correspond plus ni à la réalité du métier ni au besoin des institutions. Ainsi les archivistes sont souvent placés au sein de services extrêmement divers allant des directions des systèmes d'information (DSI) aux laboratoires de recherche en passant par les directions juridiques.

Conclusion

Le métier d'archiviste recouvre des réalités très différentes, à tel point que nous pourrions parler des métiers d'archivistes. De plus en plus, l'archivistique fait désormais partie intégrante du travail de recherche, comme en témoigne l'émergence de thèses en archivistique à l'université d'Angers. Pratiquée à des fins de mise à disposition, l'archivistique est soumise à une double tension exercée d'une part par les technologies de l'informatique et, d'autre part, par des problématiques scientifiques liées à la réutilisation de l'information. La mise en scène des archives sur Internet ne doit pas pour autant effacer la traçabilité des différents états des documents. Ainsi, l'archiviste, tour à tour webmestre, chef de projet et spécialiste en conservation des documents papiers et numériques apporte son expertise pour la valorisation scientifique d'archives tout en menant à bien sa mission de service public.

Sarah CADOREL
Archiviste/ingénieure d'études
Centre de données socio-politiques (UMS Sciences Po/CNRS)
sarah.cadorel@sciencespo.fr