

Atelier n°1. Comment décrire les archives à l'heure du web ?

35 participants

La décision de proposer un atelier sur le thème de la description et de l'indexation archivistiques à l'heure du web constitue une réponse à une demande exprimée par les représentants du réseau territorial dans le cadre du comité de suivi du portail Francearchives. Cette demande a été confirmée par le sondage que le bureau de l'accès aux archives et de la valorisation numérique a lancé début juillet sur les pratiques d'indexation, dont les réponses ont largement été exploitées dans le cadre de la préparation de cet atelier.

Isabelle Homer (directrice des Archives départementales de la Marne) a ensuite présenté l'enquête réalisée par l'Association des Archivistes français sur les sites internet des Archives départementales depuis décembre 2018. Lors d'une première phase, les sites internet ont été examinés grâce à une grille de 63 questions. Actuellement, une enquête est en ligne en partenariat avec la *Revue française de généalogie* et des ateliers tests sont conduits dans quinze services. Les conclusions encore provisoires sur l'hétérogénéité des stratégies de recherche, la mise en valeur des fonds numérisés et les difficultés de compréhension des usagers seront testées et consolidées dans la suite de l'enquête. Les résultats définitifs seront exploités lors des RASAD organisées les 6 et 7 février 2020.

Louis Vignaud (Archives nationales) a montré la nécessité pour l'archiviste de maîtriser la description archivistique presque dès sa création par le producteur. En effet, ce dernier produit des documents qu'il transfère ensuite à un service d'archives. À ce moment-là, l'archiviste doit collecter également suffisamment de métadonnées qui décrivent les documents. Ces dernières sont stockées dans une base de données et exploitées par des machines pour proposer une restitution à des utilisateurs autant humains que machines. La machine constitue un utilisateur à ne pas négliger puisque c'est elle qui va opérer une première médiation des données. Les métadonnées sont ainsi converties pour être partagées sur les réseaux sociaux ou récupérées par des portails agrégateurs (du type de FranceArchives) ou les moteurs de recherches généralistes du web. Enfin, elles sont mises à disposition des utilisateurs humains qui, parfois, ont besoin de la médiation d'un archiviste pour y accéder.

Les participants ont ensuite été répartis en trois groupes pour réfléchir à la description, à l'indexation et à l'utilisation de référentiels. L'objectif de ces sous-groupes était de déboucher sur des propositions, des suggestions concrètes, mais également d'identifier des problèmes à résoudre et des attentes à combler.

Groupe description

L'objectif était ici d'analyser nos pratiques de description en partant d'une réflexion autour des avantages et inconvénients/freins de la description XML-EAD (la plus répandue aujourd'hui).

La description en XML-EAD constitue une pratique métier identifiée et reconnue, permettant la coopération des agents autour d'une description encadrée, structurée. La structuration de l'EAD permet de créer de l'interopérabilité, facilite la transmission de données dans le cadre de projets communs à travers l'utilisation du protocole OAI-PMH. Les éléments de la description sont répartis et structurés au sein de balises, ce qui permet de différencier/classer les différents éléments d'information. Cette structuration peut être exploitée dans le paramétrage d'un moteur de recherche, mais aussi à l'affichage.

L'EAD constitue un format d'édition qui anticipe une lecture linéaire, relativement exhaustive, opérée par un humain. Ce format de description semble hésiter entre deux états et fonctions : document ou donnée ? La souplesse de l'EAD (seuls 8 éléments sur 146 sont obligatoires) et les choix pluriels d'affichage associés à la tentation de produire la donnée avant tout en fonction de ce paramètre n'ont pas favorisé la création de métadonnées homogènes. Côté utilisateur, le cadre arborescent de la description n'est pas toujours bien compris, l'information apparaît dispersée entre les différents niveaux, les termes employés sont perçus comme jargonnant. Côté archivistes, la complexité de la description hiérarchique et le nombre de balises représentent un frein certain pour une prise en main par l'ensemble des agents d'un service, et la description se révèle souvent chronophage.

Dans un second temps le groupe a été invité à proposer des actions pour mieux tirer profit des avantages et lever les freins. Les actions citées ont été rassemblées autour de deux grands axes :

1. Adopter une description orientée utilisateur

Au moment de la conception d'un instrument de recherche, l'archiviste doit veiller à adapter le niveau et la qualité des métadonnées en fonction des besoins de l'utilisateur. Plusieurs enquêtes (notamment : Association des archivistes français, FranceArchives) sont en cours ou en projet, dont les résultats pourraient être mutualisés afin de permettre aux archivistes de mieux comprendre les besoins et les usages des utilisateurs. Ces derniers pourraient également être associés à l'enrichissement des métadonnées archivistiques grâce au développement de pratiques collaboratives (indexation, description, transcription).

Les points d'accès (agents, lieux, sujets) sont autant d'entités qui interagissent, gravitent autour, conduisent vers les documents et doivent être envisagés en complémentarité de la description. Il convient également de renforcer la coopération autour de référentiels, les échanges de données d'autorité dans le réseau archives, en lien avec celui des bibliothèques. Cela permettrait d'offrir aux utilisateurs de nouveaux moyens d'accès aux métadonnées archivistiques, en particulier par une carte.

2. Production archivistique : passer du document à la donnée

Adopter une grammaire EAD commune, définir un socle commun de description, augmenter l'interopérabilité des données dans la perspective qu'offre le nouveau cadre conceptuel Record in Context (RiC), telles sont les principaux axes de travail définis par le groupe. Une démarche d'amélioration de la qualité des métadonnées archivistiques doit être menée conjointement par le SIAF, les services d'archives et les éditeurs de logiciels. La constitution d'un groupe de travail mixte piloté par le SIAF permettra d'élaborer un guide de pratiques communes pour les archivistes et de définir un socle minimal de fonctionnalités à demander aux éditeurs de logiciels pour faciliter la description. La définition de trames de description et d'indexation pour des séries documentaires très homogènes telles que les délibérations communales, les permis de construire, l'enregistrement par exemple, constituerait un réservoir de descriptions permettrait de proposer un corpus de descriptions plus homogènes. Par ailleurs, cette réflexion peut s'inspirer de la problématique de la description des archives nativement numériques. Le cadre normatif doit également évoluer : le SIAF pourrait assurer l'interface entre les organismes de normalisation et le réseau des archives, diffuser des informations régulières sur l'avancement des travaux et fournir un accompagnement pour envisager la transition vers RiC.

Enfin, un important besoin en formation directe des agents et d'outils pour acculturer les services à une description orientée données, a été pointé.

Groupe indexation

Une série d'exercices initiaux a permis d'établir un constat sur les pratiques d'indexation dans les services.

Sur les 8 participants, tous les services indexaient. Dans certains cas, les agents ne participaient pas à l'indexation, le problème relevé étant un problème de qualification ; dans d'autres cas, les agents indexaient avec une phase de relecture/validation du directeur et/ou de l'adjoint. Le cas particulier de l'indexation collaborative a été abordé très succinctement avec la question du contrôle de cette indexation et de la cohérence avec celles faites par les archivistes.

Il est ressorti de la discussion que ce qui était indexé systématiquement et semblait le plus important était les lieux, cette indexation étant la plus utile au regard des demandes des usagers, qui sont notamment demandeurs de visualisation cartographique, et permettant des recherches trans-collections (archives et bibliothèque). Ont été relevées des questions concernant les modifications territoriales (problème des circonscriptions d'Ancien Régime, fusions de communes, etc.) et un besoin de savoir quelles bonnes pratiques adoptées vis-à-vis notamment de ces modifications dans le temps. Pour l'indexation des lieux, l'utilisation de référentiels locaux était majoritaire, certains services s'appuyant sur Système d'Information Géographique. Lorsque l'indexation se fait au niveau de la rue, il a été précisé que les normes n'étaient pas suivies, afin de pouvoir obtenir un classement alphabétique des rues, très utile pour les usagers.

Les autres types d'indexation (matières / thématiques ; personnes – physiques ou morales ; typologies) apparaissent secondaires après les lieux.

Pour l'indexation matière, c'est l'ex-thésaurus W qui est majoritairement utilisé. Les participants qui l'utilisent font remarquer qu'il n'est pas parfait mais qu'il a le mérite d'exister. Il permet en utilisant les mêmes termes sur des documents de différentes époques de voir l'évolution de la société. Les services ont fait remonter le besoin de mise à jour du thésaurus, et l'importance pour eux que les éditeurs implémentent correctement cet outil afin qu'ils puissent l'utiliser dans toute sa richesse (hiérarchie et « employé pour » par exemple). Pour les documents iconographiques, l'emploi de thésaurus spécialisés a été plusieurs fois mentionné.

Pour l'indexation des personnes, plusieurs participants ont indiqué que l'indexation des personnes morales n'était pas aisée et qu'avoir un référentiel externe auquel se référer pourrait les aider. Certains se posaient également la question de définir qui devait être indexé : des personnalités ou aussi des inconnus ? De la même manière, la question de savoir pour qui on indexait a également été posée. L'indexation du réalisateur d'un film qui a été projeté lors d'un événement peut être frustrant pour l'utilisateur qui s'attend plutôt à des documents sur le réalisateur en question (photographie, correspondance, interview etc.) mais, d'un autre côté, ce genre d'indexation permet de dresser la liste de toutes les représentations des symphonies d'un compositeur dans un lieu donné, par exemple.

L'indexation des typologies sert autant aux internautes qu'aux services d'archives, à qui cela permet d'associer les typologies à des formulaires de recherche ou des modes de recherche particulier (recherche alphabétique sur certaines typologies). Les typologies de documents (carte postale, plan, acte notarié, etc.) répondent notamment au besoin de facettes exprimé par les usagers.

En conclusion, les participants s'accordaient à dire qu'un guide des bonnes pratiques les aiderait, notamment pour la formation et la montée en compétence de leurs équipes. Très souvent, des guides existent déjà en interne, et des échanges de bonnes pratiques ont lieu lors de réunions au sein des

services, mais un guide piloté et validé par le SIAF permettrait d'homogénéiser les pratiques entre services et de trancher quand des approches différentes surgissent au sein d'un même service.

Groupe référentiels

Comment tirer le meilleur parti des référentiels ? Bien que pléthorique, l'offre ne correspond pas dans les faits à une utilisation systématique des différentes listes d'autorité, thésaurus et normes dédiés à la description archivistique. Ce sous-atelier consacré aux référentiels se proposait de recueillir l'expérience des différents participants puis de dessiner les pistes d'une feuille de route commune pour une meilleure exploitation de l'existant.

Première partie : recueil d'expérience

Un premier temps a été dédié à recenser les différents référentiels utilisés par les participants et à identifier les différents freins à leur utilisation.

Champs couverts	Référentiels les plus utilisés	Points d'attention
Matière	Thésaurus pour l'indexation des archives locales (le thésaurus-matière et ses listes d'autorité), RAMEAU, index local	Il a été souligné des manques souvent à combler en matière d'iconographie (là où RAMEAU prend le relais du thésaurus anciennement W par exemple)
Lieux	Code officiel géographique (INSEE), index géographique local	Les retours des participants indiquent que le niveau de granularité des référentiels de lieux (comme Geonames) est souvent inadapté à leurs besoins
Personnes	Norme ISAAR (CPF), ISNI, index local	

Parmi les freins mentionnés à l'utilisation des référentiels, on trouve :

- Leur méconnaissance ;
- Leur inadaptation aux besoins des services (granularité, spécificités thématiques) ;
- La difficulté de les implémenter aux systèmes d'information.

Participer à un référentiel commun d'entités ?

Alors que le projet d'un Fichier national d'entités se dessine plus clairement pour l'année 2020, les participants ont été invités à réfléchir aux conditions qui permettraient de contribuer à un tel référentiel.

Le Fichier national d'entités (FNE) est un projet mené conjointement par la BnF et l'ABES visant à créer un fichier regroupant dans une même base de données toutes les entités (personnes, collectivités, lieux, concepts etc.) produites en France par les organismes dépositaires de collections documentaires. Cette base permettra aux producteurs de données d'y gérer et d'y produire

directement leurs entités. Les objectifs sont de mieux intégrer les référentiels dans le web de données, décloisonner et faciliter leur réutilisation, mutualiser la production, améliorer la qualité et la fiabilité des données et enfin mettre en avant leur complémentarité. Après une preuve de concept sur l'outil Wikibase, le FNE sera lancé en janvier 2020. Il y aura un appel au réseau afin de perfectionner l'outil selon les besoins. Bien que ce projet soit initié par le monde des bibliothèques, les autres producteurs de données que sont les musées et les archives ont tout intérêt à s'y intégrer et sont à ce titre conviés à s'associer aux différents groupes qui seront mis en place.

Au sein de l'atelier, les discussions ont porté sur l'intérêt de contribuer à un référentiel commun national. Il en ressort qu'il serait particulièrement indiqué pour l'identification des producteurs, de personnes (type « gloires locales ») et des lieux. Les avantages principaux seraient :

- Un gain considérable en termes d'efficacité au moment de la saisie lorsqu'une entité a déjà été intégrée dans la base par un autre établissement ;
- La garantie de la qualité et de l'interopérabilité des données, automatiquement mises à jour, notamment grâce à l'utilisation d'identifiants pérennes comme les ark ou les ISNI ;
- La réponse à un besoin identifié. Un fichier d'entités, mutualisé à l'échelle nationale permettrait de prendre le relais d'initiatives telles que le projet ICA-AtoM sur les notices de producteurs, aujourd'hui inactif.

Cependant, une contribution étendue pose la question de l'homogénéité des données. On peut s'attendre également à un travail de reprise ou d'enrichissement des données pour des notices créées par des producteurs n'ayant par exemple pas les mêmes exigences en termes d'informations.

Conclusion : des pistes pour une meilleure utilisation des référentiels

Les échanges ont mis en évidence un besoin de faire évoluer l'existant. Un référentiel comme le Thésaurus pour la description et l'indexation des archives locales (anciennement thésaurus W), parmi les plus utilisés par les participants, bien loin d'être en déshérence, mériterait de voir son accès et son utilisation simplifiés. La priorité porte également sur l'établissement d'un guide de bonnes pratiques pour l'indexation, un constat émis dans les autres sous-ateliers.

Il y a en général une attente d'information via la mise à disposition d'un panorama de l'offre et d'une mise à jour des circulaires. Le SIAF doit d'autre part servir d'interface de dialogue avec des acteurs comme les bibliothèques (pour le projet FNE) et les éditeurs de logiciels (pour une meilleure intégration des référentiels dans les systèmes).

Conclusions de l'atelier

Les travaux des trois groupes et la synthèse collective qui les a suivis ont permis de dégager collectivement des actions prioritaires à lancer :

- Adopter une approche orientée utilisateurs en mettant en commun les résultats des nombreuses enquêtes en cours (AAF, Francearchives) ;
- Lancer une réflexion sur l'évolution de la description archivistique et de l'indexation en prenant en compte l'arrivée de RiC et des archives nativement numériques ;
- Redéfinir les enjeux et objectifs de l'indexation et, en fonction du résultat, lancer une révision du thésaurus pour l'indexation des archives locales ;
- Lever les obstacles techniques, grâce à un dialogue concerté SIAF/réseau avec les éditeurs de logiciels ;

- Lever les obstacles RH, grâce à la formation et l'accompagnement des équipes, et notamment par le biais de l'élaboration d'un guide d'indexation.

La question – à plus long terme – de la capacité à gérer automatiquement certains aspects des multiples transitions en cours afin d'en simplifier l'appropriation par les services d'archives a également été évoquée. Le modèle conceptuel et l'ontologie Record in Context (RiC) doivent être publiés par l'EGAD en novembre 2020. Les Archives nationales sont en train de finaliser un logiciel permettant la conversion des métadonnées archivistiques aux formats EAD et EAC en jeux de données RDF conformes à RiC-O. Le code source de ce logiciel sera disponible en open source et la conversion des instruments de recherche des Archives nationales donne des résultats encourageants.

Concernant la reprise de l'arriéré du stock d'instruments de recherche, qui devra prendre en compte les évolutions en matière de description et d'indexation, il est prévu d'étudier dans quelle mesure elle pourrait être automatisée, par exemple par l'intermédiaire de francearchives.fr. Par ailleurs, les Archives nationales mènent actuellement, avec une équipe de l'INRIA, un projet de recherche de fouille des instruments de recherche pour en extraire des entités nommées.