

Dans le cadre du programme d'acquisition de ressources numériques sous licence nationale portées par le GIS CollEx-Persée, l'Abes a acquis le corpus des archives parlementaires britanniques du XIX<sup>e</sup> siècle, et choisi d'en assurer le signalement à l'aide de Calames.

# Les archives du XIX<sup>e</sup> siècle du Parlement britannique dans Calames

Fin 2020, soutenu par un financement du GIS CollEx-Persée, l'Abes a procédé à l'acquisition sous licence nationale du corpus des archives parlementaires britanniques du XIX<sup>e</sup> siècle<sup>1</sup> numérisées par ProQuest et disponible via la base *Nineteenth-Century Parliamentary Papers*<sup>2</sup>. L'intérêt scientifique de ce corpus est considérable : regroupant plus de 80 700 documents parlementaires produits par la Chambre des Communes (*House of Commons*) entre 1801 et 1900, il constitue une source primaire incontournable pour les chercheurs en études anglaises, droit, sciences politiques ou histoire contemporaine.

## LE CHOIX DE CALAMES

Une fois le corpus acquis, en coordination avec le groupe de travail CollEx-Persée en charge des acquisitions en licence nationale, l'Abes a décidé de procéder à son signalement en s'appuyant sur Calames, catalogue des archives et manuscrits de l'ESR, la structure EAD (*Encoded Archival Description*) s'avérant la plus adaptée au traitement de ce corpus. En effet, la force de Calames est de conjuguer principe archivistique de respect des fonds – décrits en arborescence grâce au format EAD – et possibilités de recherche étendue, sur chaque métadonnée, chaque document, par mots clés et par index<sup>3</sup>.

D'ordinaire, et ce depuis près de dix ans, l'Abes utilise les formats (RDF) et technologies du Web sémantique pour modéliser et intégrer les données fournies par les éditeurs et autres fournisseurs. Pour procéder à ce traitement spécifique réalisé pour la première fois par l'Abes, il convenait donc de concevoir un processus technique permettant la conversion, l'enrichissement et la production en masse de données EAD.

## DE L'INDEXATION À LA CONVERSION DES DONNÉES EN EAD

Dès les années 1980, une équipe britannique avait procédé à une première indexation du corpus, un travail titanesque, né -déjà- du constat d'un manque d'appropriation par la communauté universitaire, démunie face à une telle mine d'informations. Cette indexation originale, qui a fait l'objet d'une conver-

sion en langage d'indexation Rameau, a largement guidé l'équipe confrontée à la nécessité de classer et ordonner ces 80 755 documents d'archives, une opération indispensable en préalable à leur traitement catalographique. Pour ce faire, l'Abes s'est appuyée sur les typologies du fonds, témoins de circuits législatifs distincts, les descriptions d'archives ayant ensuite été classées chronologiquement à l'intérieur de chaque typologie.

Afin de parvenir à la structure EAD attendue par Calames, il a ensuite fallu fusionner trois sources d'informations, soit trois lots de données livrés par le fournisseur : un copieux fichier XML au typage des données peu précis et comportant beaucoup d'informations en <note> ; des fichiers MARC ; un fichier Excel comportant la liste des documents acquis, liste faisant foi d'un point de vue contractuel. Ces trois sources ont ensuite été chargées dans une base XML. Grâce aux performances et à la fiabilité du logiciel BaseX et à la puissance du langage XQUERY, ces données hétérogènes ont pu, requête après requête, être mixées puis injectées dans la structure hiérarchique de l'EAD.

En prime, lors des opérations de conversion entre l'indexation originale et l'indexation Rameau, les correspondances ont été établies pour les concepts les plus fréquents, et regroupées dans un fichier Excel qui a complété les trois autres sources.

Ainsi, en cohérence avec le programme CollEx-Persée et conformément à sa mission de mise à disposition des ressources acquises sous licence nationale, l'Abes investit le champs du signalement des archives numérisées (non localisées dans un établissement du réseau), des archives nativement numériques et des données de la recherche en portant une attention toujours plus marquée aux besoins des chercheurs et à la valorisation de leurs travaux.

ÉTIENNE NADDEO

Responsable du réseau Calames  
Département des Métadonnées et services aux  
réseaux de l'Abes  
naddeo@abes.fr

YANN NICOLAS

Responsable du Labo de l'Abes  
nicolas@abes.fr

[1] <https://www.licencesnationales.fr/les-corpus-acquis/archives-du-parlement-britannique-du-xixe-siecle>

[2] <https://parlipapers.proquest.com/parlipapers/search/basic/hcpbasicsearch>

[3] Pour ce fonds précis, les index de lieu, de collectivité, de titre d'œuvre, de date, de provenance et de langue seront utilisés.