



Bilan scientifique Octobre 2022

Coordinateur scientifique de la filière IST : Cécile Formaglio, Conservatoire national des arts et métiers

Coordinateur scientifique chercheur : Michel Lallement, Conservatoire national des arts et métiers

FamiliLettres est le projet d'édition numérique des correspondances privées, actives et passives, de Jean-Baptiste André Godin (1817-1888), industriel et créateur de l'expérimentation sociale inspirée du fouriérisme que fut le Familistère de Guise, et de sa collaboratrice et compagne Marie Moret (1840-1908). Le corpus édité sur la plateforme EMAN, particulièrement volumineux, conservé au Cnam et au Familistère, comprend environ 20 000 lettres (18 000 lettres de correspondance active, 1 950 lettres de correspondance passive). Le projet comportait aussi un volet expérimental sur la visualisation du corpus. Il s'est déroulé de septembre 2020 à mi-octobre 2022.

Le projet a été porté par la Direction des bibliothèques et de la documentation du Cnam et le Familistère de Guise avec deux laboratoires du Cnam, le Lise (Laboratoire interdisciplinaire pour la sociologie économique, UMR CNRS-Cnam n° 3320) et le Cédric (Centre d'études et de recherche en informatique et communications, EA4629). Il a aussi été porté par [EMAN](#) (Édition de Manuscrits et d'Archives Numériques), gérée au sein du laboratoire Thalim en partenariat avec l'Institut des textes et manuscrits modernes (ITEM), qui est une plate-forme de publication numérique pour la diffusion et l'exploitation de documents et de fonds d'archives qui repose sur le logiciel Omeka à partir duquel elle développe des extensions pour l'édition numérique savante.

Équipe projet

L'équipe projet qui comptait des représentants de tous les partenaires impliqués comprenait deux enseignants chercheurs du Lise (Michel Lallement) et du Cédric (Pierre Cubaud), deux conservateurs (Cécile Formaglio, Cnam et Frédéric Panni, Familistère), deux chargées de collections (Florence Desnoyers, Cnam et Marine Baron Le Canderff, Familistère), le responsable de la plateforme EMAN (Richard Walter) ainsi qu'un ingénieur d'études (Alexandre Wauthier) recruté grâce au financement CollEx-Persée entre janvier 2021 et octobre 2022. Elle s'est renforcée durant six mois (1/02/2021 au 23/07/2021) d'une stagiaire de Master 2 en design d'interaction, design graphique et visualisation de données (Laura Swietlicki).

Elle s'est réunie neuf fois en formation plénière au cours du projet et de nombreuses autres fois en formation restreinte pour des réunions d'instruction sur le schéma de métadonnées notamment et le protocole de transcription.

Notre équipe a par ailleurs pu profiter des nombreuses occasions d'échanges avec les responsables des autres projets hébergés sur EMAN (journées internes, journées publiques, workshop) et des formations et ateliers proposés par le responsable de la plateforme (formation à l'édition sur EMAN, formation à la TEI et à l'outil Transcript, ateliers par projet).

L'avancement du projet a aussi pu bénéficier de l'apport du travail de plusieurs collègues de la bibliothèque du Cnam, à la faveur notamment de la mise en place du travail à distance lié à la crise sanitaire. Deux prestataires de numérisation (Flashcopy) et développements informatiques (Numerizen) ont aussi été sollicités dans le cadre du projet.

L'équipe du projet a pu s'appuyer enfin sur un conseil scientifique présidé par Michel Lallement et composé de Bernard Desmars, maître de conférences en histoire contemporaine à l'Université de Lorraine, Jessica Dos Santos, docteure en histoire contemporaine, chercheuse associée à l'IRHIS (UMR 8529) et Rebecca Rogers, professeure d'histoire de l'éducation à l'Université Paris-Cité. Ce conseil scientifique s'est réuni à deux reprises et ses membres ont aussi été sollicités pour participer à l'étude sur la visualisation du corpus menée par Pierre Cubaud. Plusieurs d'entre eux, enfin, ont participé en tant qu'intervenant·e·s à des table-rondes à la journée d'étude d'octobre 2022 (*cf. infra*).

Objectifs du projet

- Préserver et faciliter l'accès à un fonds dispersé, peu lisible, difficile à manier
- Accompagner chaque lettre d'une numérisation de très bonne qualité
- Enrichir les métadonnées descriptives de premier niveau (titre, date, collation) avec une indexation sujet précise et une géolocalisation
- Transcrire une sélection pertinente d'items
- Enrichir l'édition des lettres avec la documentation contemporaine disponible notamment au Cnam et au Familistère : journal du Familistère *Le Devoir*, etc.
- Identifier et qualifier les correspondants de Godin et Moret
- Établir les liens avec les dictionnaires biographiques en ligne ou les bases numériques existantes
- Faire évoluer EMAN pour faciliter l'interrogation d'un corpus volumineux
- Expérimenter les méthodes de visualisation d'un corpus volumineux

Les principes d'éditorialisation de FamiliLettres s'appuient largement sur les recommandations contenues dans le [Guide méthodologique pour l'édition numérique de correspondance](#) établi par le groupe de travail « Correspondance » CAHIER (Huma-Num2016-2018) ainsi que sur les pratiques des autres projets d'édition de correspondances hébergés sur EMAN ou ailleurs. Nous avons pu aussi bénéficier des retours du Groupe de travail "Correspondance" organisé au sein d'EMAN.

Si notre [corpus](#) intéresse de nombreux champs de la recherche historique (histoire du fouriérisme et du socialisme, de l'éducation, féminisme, de l'architecture du logement, du pacifisme, histoire du mouvement spirite...), et des sciences sociales en général, nous avons essayé de ne pas préjuger par nos choix d'édition des usages qui pourraient en être faits et d'ouvrir au maximum le champ des possibles à l'exploitation scientifique ou amateur. Les difficultés posées par la navigation et l'interrogation d'un corpus volumineux et notre volonté de tenter d'y répondre ont particulièrement guidé les choix qui ont été faits.

Méthode

◆ **Projet éditorial:**

Les premiers mois du projet ont été consacrés à nous former à l'outil EMAN, à élaborer une méthode de travail avec son responsable pour l'alimentation de notre site et à réfléchir à notre schéma de métadonnées, à travers le traitement test d'un échantillon de six registres de lettres de la correspondance active. Il a été décidé de travailler par import des données sous forme de tableurs, plutôt que directement dans EMAN, ce qui facilitait la relecture avant import et les modifications en masse sur les données, dans la mesure où nous travaillions dans un espace partagé en ligne, accessible à tous les membres de l'équipe projet.

La structuration en collections imposée par OmeKa nous a amené-e-s à décider d'une organisation séparant les métadonnées liées aux lettres et celles liées aux correspondant.e.s, dans le but notamment d'éviter les saisies redondantes.

Enfin, il s'est agi de déterminer la liste des métadonnées descriptives pour les lettres (Dublin Core ou métadonnées personnalisées) et d'élaborer pour chacune des consignes de saisie. Il a par ailleurs été décidé assez rapidement de joindre à chaque lettre un résumé de son contenu afin de rendre plus facile la consultation du site.

De nombreuses réunions ont été organisées pour l'élaboration de plusieurs listes contrôlées: sujet, événements cités, œuvres citées notamment. Pour les mots sujet, Rameau étant apparu rapidement comme inadapté à nos besoins et au niveau de granularité descriptive souhaité, tout l'enjeu était de trouver le bon niveau tout en évitant de créer une liste trop complexe ou trop longue et prenant en compte l'historicité des mots. Pour les œuvres citées, nous nous sommes appuyé-e-s sur une bibliographie collective dans Zotero, alimentée au fur et à mesure du repérage des références mentionnées dans les lettres lors de l'édition de la correspondance, et éditée grâce à un style

personnalisé pour notre projet, pour nous assurer de l'homogénéité des données du champ « œuvres citées ».

Pour les correspondant.e.s, des consignes de rédaction ont été rédigées pour la forme à donner à la notice biographique. Nous nous sommes évidemment appuyés sur les réservoirs d'autorités (BnF et IdRef). Les notices correspondant.e.s. ont fait l'objet d'une mise à jour constante jusqu'à la fin du projet, et vont continuer à être enrichies, soit grâce aux informations provenant du traitement de la correspondance, soit grâce aux recherches biographiques menées. Des métadonnées requêtables ont complété les biographies: le genre, le pays d'origine (plutôt que la nationalité, information jugée trop mouvante au 19^{ème} siècle). Une réflexion, soumise en particulier à notre conseil scientifique à deux reprises, menée jusque dans les derniers mois du projet, a porté sur la qualification de l'activité des correspondant.e.s pour arriver à l'élaboration d'une autre liste de mots contrôlés décrivant l'activité professionnelle, l'activité intellectuelle et le positionnement dans la hiérarchie sociale. Le parti-pris a été d'essayer de n'exclure aucun champ et de ne pas mettre en avant les activités de premier intérêt pour notre corpus (fouririsme ou socialisme par exemple) au détriment par exemple des activités commerciales pour ne pas restreindre les usages du site. Les référentiels socio-professionnels usités au 19^{ème} siècle pour les recensements et ceux élaborés par les chercheurs travaillant sur la période ont ainsi été compulsés pour arriver à l'élaboration du nôtre.

Le principe d'une relecture systématique a été adoptée pour tous les contenus du site et le travail d'enrichissement des métadonnées a été réparti entre les membres du projet.

Afin de constituer le corpus définitif, une réflexion était nécessaire sur la notion de correspondance et sur les types de documents à retenir dans le cadre du présent projet. Ont ainsi été considérées comme des lettres tous les documents écrits, signés et adressés à une personne physique, à un groupe de personnes ou à une personne morale, sans que cet échange ait forcément fait l'objet d'un envoi postal. Ont ainsi été inclus par exemple dans le corpus des billets, des cartes postales, des cartes de visite, des factures. Les registres de copies de correspondance active ayant fait l'objet d'une numérisation avant le lancement du projet, ces typologies ont été appliquées au repérage systématique dans les archives du Cnam et du Familistère de la correspondance active et passive non numérisée. Les six premiers mois de l'année 2021 ont été consacrés à ce repérage et à la préparation et à la description des 2293 lettres repérées et envoyées en numérisation chez notre prestataire Flashcopy en juin 2021.

Enfin, une importante refonte de l'outil de Transcription, Transcript, proposé par EMAN nous a amenés à décaler notre réflexion sur la transcription qui est intervenue dans la dernière année du projet. Des principes de transcription, beaucoup débattus, ont été élaborés ainsi qu'un choix de balises

TEI à utiliser. [Ces principes](#) pour une transcription linéaire enrichie recherchent un équilibre entre le respect de la graphie des auteurs et un souci de lisibilité.

◆ Expérimentation sur la visualisation des données

L'étude sur la visualisation du corpus a commencé par un état de l'art et le repérage des références bibliographiques traitant de la visualisation du patrimoine culturel. 166 références traitant de correspondance ont été compilées et analysées.

Deux ateliers de co-conception ont ensuite été organisés. Le premier atelier, qui s'est tenu en ligne le 19 mai 2021, et qui a réuni huit personnes associées au projet, a consisté en l'analyse de huit sites internet sélectionnés car représentatifs de l'état de l'art. Nous avons utilisé l'outil de tableau blanc Miro. Le but de ce premier atelier était le repérage d'outils utiles à la visualisation globale que nous souhaitons pouvoir offrir pour notre corpus.

Le travail de design mené par Pierre Cubaud et sa stagiaire a ensuite été soumis à l'équipe projet lors d'un workshop organisé le 11 octobre 2021 au Familistère de Guise et rediscuté collectivement. Ils ont proposé une représentation nouvelle appelée "matrice de correspondance", une interface destinée à être zoomable et qui procure une vue globale du corpus selon les correspondants et la chronologie des échanges. La matrice interagit avec des représentations cumulatives (carte, nuage de mots, graphiques barres) pour permettre des recherches sélectives dans le corpus. Cette matrice est appelée à encore évoluer pour prendre en compte l'avancée du traitement du corpus.

◆ Questions juridiques

Le droit moral relatif à la divulgation d'œuvres inédites s'appliquant à la correspondance nous avons sollicité l'autorisation auprès des ayants droit de Marie Moret, avec qui le Familistère était en contact pour la correspondance active. Le problème étant plus complexe pour la correspondance passive, regroupant plusieurs centaines d'auteurs, nous avons consulté le répertoire Balzac (Répertoire des auteurs et de leurs ayants droit) créé par la Société des gens de lettres (SGDL) pour 309 auteurs dont nous connaissons les dates de vie et de mort. Sur les 163 auteurs présents dans la base, aucune coordonnée d'ayants-droits n'était connue de la SGDL.

Nous envisageons à présent de publier la liste sur internet de ces auteurs afin d'essayer de solliciter les ayants droit à se manifester auprès de nous, afin d'obtenir leur autorisation. Néanmoins, à ce stade, nous envisageons la mise en ligne de la correspondance passive, même en cas de non-identification.

Enfin, l'édition de la correspondance étant aussi concernée par le respect à la vie privée ou tout le moins par l'attention à ne pas porter atteinte à l'honneur des personnes, il a été convenu avec notre conseil scientifique que toute lettre susceptible de poser ce genre de problème leur serait soumise pour avis.

Résultats

◆ Éditoriaux

Le [site FamiliLettres](#) a ouvert sur EMAN le 1er octobre avec de nombreuses pages de présentation du projet de nombreux contenus. 5 966 lettres de la correspondance active de Jean-Baptiste André Godin et Marie Moret ont été éditées et mises en ligne selon le schéma très complet qui a été élaboré, ce qui correspond à un tiers du corpus de la correspondance active. L'édition de plus de 4 000 lettres a été réalisée et est en cours de relecture ; elles pourront être mises en ligne rapidement.

3 306 notices biographiques de correspondant.e.s ont également été mises en ligne à la même date (3339 au 7 novembre 2022). Ces notices répertorient aussi bien les destinataires de lettres que les auteurs de lettres de la correspondance passive. La rédaction de ses notices a pu s'appuyer sur des outils existants comme [le livre des visiteurs du Familistère](#), les répertoires nationaux et internationaux d'autorités et ont donné lieu à des recherches dans l'état-civil, dans des bases ou des dictionnaires biographiques existants ou encore dans Gallica, et en particulier dans la presse et les almanachs du commerce numérisés. Les résultats de ces recherches ont ensuite été remis à disposition de la communauté en étant réinjectés dans IdRef: 39 autorités personnes physiques ont été créées, 29 notices complétées et 5 dérivées du réservoir de la BnF ; 6 autorités personnes morales ont été créées, 1 modifiée et 3 dérivées.

◆ Informatiques

Tout au long du projet des échanges avec le responsable de la plateforme EMAN ont permis de déterminer les développements informatiques nécessaires à notre projet et facilitant le travail des équipes sur EMAN. Ces derniers sont aussi mis à disposition de tous les sites hébergés sur EMAN. Ils ont été réalisés par le prestataire Numerizen encadré par Richard Walter sur différents sujets:

- **Gestion du projet** : installation d'un outil de gestion interne de projet Redmine pour gérer les demandes informatiques et les réponses apportées.
- **Personnalisation des interfaces** : EMAN a développé un plugin UITemplates qui permet de personnaliser les contenus à afficher dans une interface classique d'Omeka. Le projet

FamiliLettres a aidé à développer une fonctionnalité importante qui permet à chaque projet de personnaliser les pages de présentation des différents plugins utilisés par le projet. Car on ne présente pas de la même façon une visualisation par graphe ou par géolocalisation ou par index quand on traite un corpus antique et un corpus de correspondance du 19^e siècle par exemple.

- **Transcription (outil Transcript)** : grâce au soutien du projet FamiliLettres, nous avons opéré une optimisation du code en vue de gérer les différentes conventions de saisie des transcriptions, avec une gestion des rendus plus fine grâce à l'outil TEI Publisher (nous utilisions précédemment un autre outil, TEI Boerplate qui n'était pas assez puissant pour gérer les différents cas de figure) ; nous avons aussi réaliser une adaptation des menus déroulants qui sont devenus dynamiques, en fonction de l'endroit où on se situe dans le texte et des autres balises déjà utilisées ; cela permet de gérer des menus à taille raisonnable et proposant uniquement des balises autorisées par le schéma TEI général.
- **Module de cartographie** : nous avons changé de système cartographique en n'utilisant plus l'API Googlemap pour préférer le système ouvert Leaflet ; nous avons aussi changer de modèle de points sur la carte pour préférer un système en étoiles qui s'ouvre avec les différentes notices partageant la même zone. Nous avons surtout développer une fonction pour trouver automatiquement les coordonnées GPS et localiser un item dont l'adresse « physique » est indiqué dans une métadonnée, que cela soit DC.Coverage ou une autre. Cette fonction permet d'automatiser une grande partie de la géolocalisation, ce qui est précieux quand on a plusieurs milliers d'item à traiter. Il faut toutefois faire une vérification sur les lieux « exotiques » trouvés (comme pour le terme « inconnu » qui sera géolocalisé sur un fleuve du Vénézuéla portant ce nom).
- **Navigation** : une fonction a été développée pour que le choix des éléments du menu de page d'accueil soit fait par l'administrateur du projet et non par l'informaticien ; nous avons aussi installé un fil d'Ariane sur toutes les pages du site, reprenant les collections, les items et les fichiers concernés avec liens pour y accéder.
- **Relations** : nous avons effectué une mise en relation à grande échelle entre lettre et correspondant par personnalisation du module existant Item Relations (plus de 11.000 relations faites automatiquement) ; nous avons développé une fonctionnalité pour pouvoir faire une intégration de contenus d'une notice dans une autre (ex. afficher la biographie du correspondant dans la description de la lettre) : cette fonctionnalité est intégrée au module UITemplates et va offrir de nombreuses possibilités à tous les projets sur EMAN.

- **Recherche** : nous avons adapté le moteur de recherche interne d’Omeka 2.7 pour des recherches croisées (ex. : toutes les lettres envoyées à une destinatrice femme). Le résultat est une page autonome qu’on peut atteindre via l’ajout d’un lien sur le menu du site ; cette page propose de cumuler différents critères de recherche en choisissant des métadonnées et les opérateurs traditionnels « contient, ne contient pas, est exactement, est vide, n’est pas vide...).
- **Liens** : nous avons développé un module pour la vérification de liens cassés ou morts à l’intérieur du site, sur tout type de contenus : « simples pages », menu d’accueil et pages d’exposition, collections, items, files.

Tous ces développements ont été faits avec l’optique de pouvoir être réutilisables pour les autres projets EMAN et, dans le cas des modules diffusés sur les plateformes GitHub/GitLab, par d’autres projets utilisant Omeka.

Valorisation du projet

Le projet et ses problématiques ont été présentés lors des [journées EMAN des 28 et 29 janvier 2021](#) lors de deux présentations, “Éditer un corpus et/ou un fonds d’archives” et “Gestion des corpus étendus”.

Un premier point d’étape a été présenté à la communauté lors du webinaire EMAN “L’épistolaire à l’ère du numérique. Entre perspectives historiques et innovations technologiques”, organisé par Marie Dupond (GAPN/ENS) et Jean-Sébastien Macke (ITEM/CNRS-ENS) en 2020 et 2021 auquel nous avons été invités à participer [le 19 mars 2021](#).

Le projet a aussi été évoqué dans un [court article](#) publié dans le numéro 104 d’*Arabesques*.

Les travaux menés par Pierre Cubaud sur la visualisation du corpus ont été présentés le mercredi 21 septembre 2022 lors de [The International Conference on Theory and Practice of Digital Libraries \(TPDL\)](#) à Padoue. Le texte de la communication est consultable [ici](#).

Les résultats du projet et les réalisations à venir ont été présentés lors d’une [journée d’études le 07 octobre 2022](#) au Conservatoire national des arts et métiers jugée unanimement comme une grande réussite. Cette journée, qui a mobilisé seize intervenant-e-s et trois modérateur-ric-e-s, a réuni 31 participant.e.s. Plusieurs moments thématiques ont scandé l’évènement : présentation du projet FamiliLettres ; mise en contexte socio-historique des conditions de production du corpus ; regards croisés sur l’intérêt scientifique que peut représenter l’édition numérique de la correspondance

Godin-Moret et d'autres correspondances de statut similaire ; retours d'expériences sur le travail d'édition numérique effectué dans FamiliLettres (conditions techniques, choix méthodologiques, enjeux de la transcription...) ; retours d'expérience et premières idées d'usage possibles du corpus par de premier-ère-s utilisateur-ric-e-s. De fait, cette journée se donnait également pour objectif de soumettre notre site à ses usagers potentiels et de dresser un panorama des usages possibles au service de la recherche. À cette fin les six historien.ne.s et sociologues invité-e-s à participer à deux tables rondes ont pu bénéficier d'un accès au site avant son ouverture et tester les différents outils et modes de navigation ainsi qu'évaluer la pertinence des métadonnées choisies. De nombreuses pistes de recherche et d'exploitation du site ont ainsi pu être proposées et discutées au cours de la journée. L'équipe FamiliLettres ne s'interdit pas enfin de proposer à l'avenir à une revue spécialisée (comme les *Cahiers Charles Fourier*) un numéro spécial basé sur l'exploitation du corpus de la correspondance qui a été éditée.

Conclusion

Le site FamiliLettres est ouvert, il commence à recevoir des visites (234 visiteurs uniques au mois d'octobre et 473 visites). L'édition des correspondances de Jean-Baptiste André Godin et de Marie Moret conservées au Cnam et au Familistère de Guise, ainsi que leur transcription va se poursuivre. Le site continuera aussi à évoluer en fonction des disponibilités des porteurs de projets. Des pistes d'élargissement du corpus sont envisageables avec l'intégration des correspondances conservées dans d'autres établissements ou la mise en relations des lettres traitant du Familistère avec les photographies contemporaines conservées au Cnam et surtout au Familistère. Nous continuerons bien évidemment aussi à participer aux discussions de la communauté réunie autour d'EMAN et le site pourra ainsi évoluer en intégrant les futures fonctionnalités à l'étude sur EMAN (commentaires et transcriptions collaboratives par exemple ou mise en œuvre de l'HTR).