**Motifs**
ISSN : 2726-0399

**5 | 2022**
**Normal, anormal, anomal**

# Lettresfamiliales.ehess.fr : reprendre et enrichir l'édition scientifique d'une correspondance en état d'obsolescence technique

**Bertrand Duménieu, Danièle Poublan, Jean-Damien Généro, Francine Filoche et Patricia Bleton**

https://lodelpreprod.univ-rennes2.fr/motifs/index.php?id=606

**DOI :** test606

**Référence électronique**
Bertrand Duménieu, Danièle Poublan, Jean-Damien Généro, Francine Filoche et Patricia Bleton, « Lettresfamiliales.ehess.fr : reprendre et enrichir l'édition scientifique d'une correspondance en état d'obsolescence technique », *Motifs* [En ligne], 5 | 2022, mis en ligne le 06 décembre 2022, consulté le 21 avril 2025. URL : https://lodelpreprod.univ-rennes2.fr/motifs/index.php?id=606

# Lettresfamiliales.ehess.fr : reprendre et enrichir l'édition scientifique d'une correspondance en état d'obsolescence technique

**Bertrand Duménieu, Danièle Poublan, Jean-Damien Généro, Francine Filoche et Patricia Bleton**

## PLAN

Constitution du corpus et première publication électronique
- Une famille de scientifiques, d'industriels et de lettrés
- Une collecte minutieuse de longue haleine
- Un corpus enrichi d'un appareil critique
- 2004-2019 : première édition numérique sur le Web avec Lodel

*Lettresfamiliales.ehess.fr* (https://lettresfamiliales.ehess.fr), un wiki sémantique d'édition et de publication
- Le choix de MediaWiki
- Contenu et catégories
- Exploration : indexation et recherche
- Téléchargement et export
- Une API pour interroger et explorer le wiki
- Un corpus structuré par un graphe de connaissances exposé sur le Web des données

Un corpus épistolaire numérique ouvert et librement réutilisable

Du wiki au dépôt de données versionné : export multi-format en masse
- Licence et crédit
- Évolutions envisagées

Les leçons et attendus d'une expérience de bricolage numérique

Résumé des jeux de données déposés

## TEXTE

*Les travaux de collecte et d'éditorialisation du corpus, de transfert vers une solution logicielle MediaWiki et de dépôt ont été entièrement réalisés au sein du collectif « Sources et données de la recherche » du Centre de recherches historiques, unité mixte de l'École des hautes études en sciences sociales (EHESS) et du Centre national de la recherche scientifique (CNRS). Ce collectif est une structure transversale qui réunit des chercheurs et des ingénieurs autour des projets numériques menés par le laboratoire.*

*L'hébergement du wiki et sa maintenance applicative sont assurés par l'EHESS.*

*Nous remercions chaleureusement Hugo Chièze, chargé de communication au Centre de recherches historiques, pour son regard expert sur l'interface graphique du wiki et ses propositions graphiques.*

1 La constitution et le traitement informatique de corpus documentaires par le biais de « grandes enquêtes collectives » était l'une des clefs de voûte du Centre de recherches historiques de l'École des hautes études en sciences sociales (Bourdelais 1995 ; Lévy-Dumoulin 2005). Dans les années 1980, une équipe de ce laboratoire a commencé à travailler sur une correspondance familiale, dans le sillage de deux programmes de recherche centrés sur la pratique de l'écriture en France – l'un sur l'alphabétisation des Français (Furet et Ozouf 1977) et l'autre sur le trafic postal au milieu du XIXe siècle (Dauphin, Lebrun-Pézerat, et Poublan 1991). Repéré par Louis Bergeron, ce fonds privé est composé de lettres échangées de la fin du XVIIIe siècle au début des années 1920 par les membres de la famille Mertzdorff-Duméril-Froissart, qui compte des industriels du textile, des savants du Muséum national d'histoire naturelle et des hauts fonctionnaires civils et militaires. Leurs échanges épistolaires traduisent leur ascension sociale en restituant l'accumulation progressive du capital intellectuel, social et économique : la lettre est mobilisée comme un instrument de sociabilité permettant d'interagir avec les proches, d'étendre le cercle des connaissances et de créer de nouveaux réseaux. La conscience de cette ascension est, avec les habitudes professionnelles des épistoliers et la stabilité des résidences, un élément déterminant de la conservation et de la transmission des lettres. L'attention familiale portée à cet héritage perdure : dans les années 1980, un descendant entreprend de publier à compte d'auteur trois monographies sur sa famille, qu'il documente avec des extraits de lettres en sa possession[1].

2 Une étude des lettres (rhétorique épistolaire, rythme d'écriture, rôle social et familial), publiée en 1995, est accompagnée de l'édition d'un échantillon de cent vingt documents (Dauphin, Lebrun-Pézerat, et Poublan 1995). Au-delà de l'échantillon de l'étude, ce travail a initié en 2004 l'ouverture d'une plateforme numérique reposant sur le logiciel Lodel. Ce dernier n'étant plus maintenu à la fin des années 2010, le Centre de recherches historiques a souhaité mettre en ligne une nouvelle édition électronique, en mobilisant le moteur de wiki libre et *open source* Semantic MediaWiki. Cet article vise à rendre compte de

ce projet de reprise, qui conjugue un impératif de sauvegarde de données mises en péril par l'arrêt de la maintenance de leur hébergeur avec une volonté d'enrichissement du corpus par l'intégration de nouveaux documents et l'ajout de nouvelles possibilités d'étude, grâce au tissage de liens sémantiques. Cette opération a été l'occasion de transformer un corpus dont la collecte a été initiée il y a près de 40 ans, pour l'ouvrir sur le Web en l'enrichissant d'un graphe de connaissances modélisant le réseau de tous les individus évoqués dans les lettres. Suivant les bonnes pratiques d'ouverture des données de la recherche, le corpus est aujourd'hui muni d'une licence de réutilisation et également déposé sur les entrepôts de données ouverts Zenodo et Didomena.

# Constitution du corpus et première publication électronique

## Une famille de scientifiques, d'industriels et de lettrés

3 Les lettres conservées s'ordonnent autour de trois figures masculines : André Marie Constant Duméril (1774-1860), professeur au Muséum national d'histoire naturelle de Paris, Charles Mertzdorff (1818-1883), manufacturier alsacien dans le textile, et Damas Froissart (1852-1923), officier et propriétaire terrien. Le premier, fils d'un juge de paix d'Amiens, est admis en 1795 au titre d'« élève de la patrie » à l'École nationale de santé nouvellement créée par la Convention. Il étudie la médecine et les sciences naturelles, et une longue carrière commence pour lui au Muséum. Il écrit régulièrement à ses parents puis à sa femme, relatant ses activités et ses projets, ses rencontres, ses difficultés et ses succès – renvoyant des échos des événements parisiens, de la Révolution au Second Empire.

4 Sa petite-fille Caroline épouse en 1858 Charles Mertzdorff, industriel du blanchiment des tissus en Alsace et maire de Vieux-Thann (Haut-Rhin) de 1865 à 1872. Ils ont deux filles, Marie et Émilie, avant que Caroline ne meure soudainement (1862). Une cellule familiale se

reconstitue autour de Charles Mertzdorff et de sa seconde épouse, Eugénie Desnoyers, issue elle aussi du monde des savants du Jardin des plantes à Paris. Après la guerre de 1870, Charles Mertzdorff choisit de rester en Alsace devenue allemande, tandis que ses filles sont élevées à Paris, dans la famille de sa seconde épouse tôt décédée – d'où une correspondance nourrie entre le père et les adolescentes. Les lettres de ces années déroulent la vie quotidienne au Jardin des plantes où ces jeunes filles accomplies fréquentent des cours privés et la Sorbonne ; dans leur famille parisienne, elles côtoient les savants de l'époque : Jules Desnoyers, Henri Milne-Edwards et son fils Alphonse Milne-Edwards.

5 Marie Mertzdorff, l'aînée des deux filles, épouse en 1880 Marcel de Fréville, conseiller à la Cour des comptes ; sa sœur Émilie épouse trois ans plus tard Damas Froissart, polytechnicien, officier d'artillerie, propriétaire terrien dans le Pas-de-Calais. Après la mort de Charles Mertzdorff en 1883, le volume de la correspondance conservée diminue. Cependant, le réseau épistolaire est encore large et vivant : il est question des enfants de l'une et de l'autre (cinq pour Marie, six pour Émilie), de nouvelles concernant parents et amis communs, mais aussi d'éducation et de voyages, de réceptions et de la vie sociale des deux sœurs et des évènements politiques.

6 Les liens avec l'Alsace sont maintenus. L'usine de blanchiment d'étoffes de Vieux-Thann, avec de nouveaux statuts, reste dans la famille : Guy de Place, l'époux d'Hélène Duméril (une autre petite-fille d'André Marie Constant Duméril), en devient gérant en 1904. Sa correspondance « d'affaires » pendant la guerre de 1914-1918 (170 lettres) se mêle à celle d'Émilie et de Damas Froissart avec l'un de leurs quatre fils, Louis – la seule qui nous est parvenue.

## Une collecte minutieuse de longue haleine

7 La correspondance de la famille Mertzdorff-Duméril-Froissart était mélangée avec des écrits de natures diverses au sein des archives familiales. En outre, une partie des lettres se trouvaient en copie dans neuf volumes manuscrits rassemblés au début du XXe siècle par les Duméril. Ces registres, restés inédits jusqu'à aujourd'hui, n'avaient pas pour objectif de réunir les lettres familiales dans un ensemble chro-

nologique cohérent : ils traduisent une volonté d'accumulation de souvenirs concernant une personne disparue et menée par un de ses proches (les souvenirs du père par son fils, ceux du mari par sa veuve).

8 Quelques lettres d'autre provenance ont été intégrées au corpus : celles éditées par Paul Triaire (1892) dans son ouvrage *Bretonneau et ses correspondants*, dont André Marie Constant Duméril fait partie. Les derniers inédits ont été localisés au Muséum national d'histoire naturelle, détenteur de vingt-sept lettres signées par André Marie Constant Duméril.

9 Le corpus n'a donc pas été constitué d'un seul tenant mais au fil des années, grâce à des contacts noués avec les représentants actuels de la famille Mertzdorff-Duméril-Froissart et aux collectes menées dans les fonds d'archives et collections publiques. Un patient travail de retranscription des lettres a été entamé bien longtemps avant l'apparition d'outils de reconnaissance automatique de l'écriture manuscrite efficaces. Ce travail se poursuit encore aujourd'hui alors que le corpus est ponctuellement alimenté de nouvelles lettres.

10 Le contenu de chaque lettre a été intégralement retranscrit à l'aide d'un logiciel de traitement de texte, en conservant à la fois le contenu de la correspondance et les éventuels en-têtes imprimés. Les mentions épigraphiques ajoutées par la suite, qui signalent des gestes multiples de collecte, de conservation et de classement par les héritiers, ont été reproduites et distinguées par la typographie. La transcription est largement imitative. La syntaxe, la ponctuation, l'usage ou non des majuscules, la marque des paragraphes ont été reproduits aussi fidèlement que possible [2]. Les ratures lisibles sont représentées par des barrés, les mots illisibles ou de lecture douteuse sont placés entre crochets. Les abréviations sont toutefois développées et placées en exposant. Enfin, des opérations de normalisation orthographique ont été effectuées afin de faciliter la recherche plein texte dans les fichiers [3]. Les noms propres ont été alignés sur la forme internationale proposée par le répertoire d'autorité de la Bibliothèque nationale de France. L'homogénéité de la transcription n'est néanmoins pas absolue, notamment au niveau des dates : le mois « Xbre » peut, par exemple, être transcrit tel quel ou de manière plus explicite (« décembre »).

## Un corpus enrichi d'un appareil critique

11 Le corpus épistolaire fourmille de noms de personnes, sans qu'il soit toujours possible de les identifier, ni même de restituer la bonne orthographe – en particulier lorsque nous ne possédons que les copies des lettres. En dépit de ces incertitudes, environ 5 000 individus cités sont dénombrés dans les quelque 3 250 lettres qui forment le corpus dans son état actuel, au printemps 2022.

12 Pour une partie de ces individus ont été constituées des minibiographies individuelles ou familiales. Ainsi, environ 1 000 biographies de personnes et de familles ont été rassemblées. Mentionnant les dates de naissance et de décès, l'éducation, le parcours professionnel, les publications éventuelles et les relations possibles avec les signataires des lettres, elles permettent de restituer le contexte social ou intellectuel propre à un milieu et une époque. Elles lient expériences singulières et histoire collective. Les biographies éclairent les réseaux relationnels, dessinent la configuration des amitiés, des connaissances, des relations de travail et signalent des accès, même indirects, au même cercle, social, économique ou intellectuel. Ainsi, la correspondance familiale se trouve en relation dynamique avec quantité d'autres figures qui appartiennent à différentes « niches » sociales, à différents espaces étroitement intriqués.

## 2004-2019 : première édition numérique sur le Web avec Lodel

13 Des outils numériques ont été mobilisés afin de rassembler les documents issus de ces ensembles divers en un seul corpus ordonné chronologiquement et pour permettre l'intégration de nouvelles pièces au fur et à mesure des découvertes. Une plateforme Web de publication électronique [4] a été ouverte dès 2004, utilisant le logiciel d'édition numérique *open source* Lodel. Ce choix, fait à rebours des supports de diffusion physiques en vogue au début des années 2000 (CD ROM, DVD), a permis de maintenir une plateforme stable pendant quinze ans, alimentée en continu de nouveaux documents.

14 L'année 2019 voit l'arrêt de la maintenance et de l'évolution de Lodel par ses développeurs. Surtout, le service Web ServOO permettant de convertir les transcriptions des lettres du format Microsoft Word en XML-TEI affichables par Lodel, est fermé. Bien qu'une nouvelle version de Lodel soit publiée et qu'un outil soit proposé en remplacement de ServOO, leur maintenance et leur développement à long terme ne semblent alors pas garantis. Dans le même temps, l'École des hautes études en sciences sociales (EHESS), qui héberge le site, abandonne progressivement Lodel.

15 La pérennité du site de publication est alors compromise à moyen terme et le service d'import de nouveaux contenus est définitivement condamné. L'inévitable transfert du corpus numérique vers une solution alternative offre l'opportunité d'une publication renouvelée suivant les bonnes pratiques d'ouverture des données de la recherche.

# *Lettresfamiliales.ehess.fr* (#tocfrom6 06-6), un wiki sémantique d'édition et de publication

## Le choix de MediaWiki

16 Le corpus des correspondances est régulièrement enrichi de nouvelles lettres et l'appareil critique est complété et affiné au fil du temps et des apports documentaires. Ce mode de construction itératif et cumulatif impose le couplage fort entre le lieu de fabrication des documents (transcriptions, notes, etc.) et celui de leur exposition sur le Web. Les gestionnaires de contenus[5] comme Lodel répondent à cette nécessité en intégrant les deux dimensions dans un seul outil.

17 Pour ce travail, nous avons opté pour une republication sous forme de wiki, appuyé sur le gestionnaire de contenus libre et *open source* MediaWiki muni de son extension Semantic MediaWiki. Bien que cette technologie ne soit pas couramment utilisée pour l'édition électronique de corpus en sciences sociales, elle offre des fonctionnalités qui nous ont paru particulièrement pertinentes.

18 Moteur des projets de la Wikimedia Foundation (Wikidata, Wikisource), la caractéristique essentielle de MediaWiki est d'avoir été conçu comme un espace d'écriture collaboratif et cumulatif dont l'organisation est construite par itérations à partir du contenu. Pour cela, MediaWiki propose le langage de balisage wikicode qui permet de structurer et de mettre en forme le texte d'une page de wiki. Le texte balisé d'une page n'est pas affiché tel quel mais est interprété par l'outil pour produire le document HTML affiché par le navigateur Web. Wikicode permet également de catégoriser une page et de structurer un wiki pas à pas, en parallèle de son alimentation en contenu, sans nécessiter une phase préalable de modélisation des données.

19 En plus d'un export RDF[6] natif, l'extension Semantic MediaWiki ajoute la capacité d'associer toute page du wiki avec un ensemble de métadonnées directement insérées dans le corps du texte sous forme d'annotations. Ces métadonnées peuvent être ensuite manipulées, regroupées, croisées, filtrées et mises en page dans le wiki à l'aide de requêtes écrites en wikicode.

20 Le contenu d'un wiki est nativement versionné, permettant de conserver la généalogie complète (à partir de l'import initial) des textes édités et donne à voir toute la complexité du travail de constitution, de structuration et de publication d'un corpus de grande taille.

21 Bien que les wikis soient conçus pour le travail collaboratif, *lettresfamiliales.ehess.fr* (https://lettresfamiliales.ehess.fr) ne permet pas aux visiteurs d'intervenir et de modifier les documents édités. Seuls les chercheurs responsables du projet en ont la possibilité, afin de maintenir ainsi la cohérence globale du corpus et de contrôler entièrement la qualité de son contenu.

## Contenu et catégories

22 La totalité des documents du corpus épistolaire et de son appareil critique est répartie en quatre catégories :

- - **Lettres** (3 250 documents). Chaque lettre correspond à une page, contenant les fac-similés, la transcription du texte original enrichi de notes et de nombreux liens hypertextes vers d'autres pages du wiki. Chaque lettre est identifiée par son titre, qui est la date d'écriture. Lorsque plusieurs lettres

sont écrites le même jour, le titre prend la forme : « Lundi 14 mars 1842 (A) », « B », etc., possiblement jusqu'à « G ». Un sous-titre mentionne la nature du message (lettre, carte), le ou la signataire, le ou la destinataire, le lien qui les unit, les lieux de départ et d'arrivée. Par exemple, « Lettre d'Eugénie Desnoyers (Paris) à son amie Caroline Duméril, épouse de Charles Mertzdorff (Vieux-Thann) ». Des notes servant à identifier les personnes mentionnées sous des formes telles que « l'oncle » ou « Marie » complètent les textes. Les personnes définies par leur fonction (« le préfet », « le dentiste ») sont dans la mesure du possible identifiées en note par leurs nom et prénom.

- **Biographies** (1 000 documents). Les minibiographies concernent aussi bien des personnes citées dans les lettres que des destinataires ou des auteurs. Les éléments biographiques explicitent une étape dans un parcours, marquent un nœud relationnel. Par exemple, une attention particulière est accordée aux médecins cités par André Marie Constant Duméril : ses professeurs, ceux qui l'hébergent lors des tournées de jury de médecine, ceux qu'il soutient ou qui le soutiennent pour l'obtention d'un poste, etc.
- **Monographies** (110 documents). Un ensemble de monographies succinctes complète la documentation. Elles donnent des précisions sur des institutions, des lieux ou des événements qui reviennent de façon récurrente dans les lettres, le plus souvent de façon allusive. Elles restituent pour le lecteur actuel les connaissances partagées par les épistoliers, ce savoir commun, factuel et affectif, que chacun devine en filigrane lorsque, par exemple, le mot « Jardin » [des plantes] apparaît sous la plume.
- **Papiers familiaux et documents divers** (50 documents). Les lettres sont conservées dans des archives familiales, riches de quantité d'autres documents. Ceux relatifs à la famille sont publiés ici : journaux personnels ou de voyages, testaments, certificats scolaires, etc.

23 Chaque document – lettre, biographie, monographie ou papier – est représenté par une page unique sur le wiki, identifiée par son titre. Les textes sont formatés en wikicode. Par conséquent, le rendu HTML des pages diffère significativement de leur texte balisé. La figure 1a présente la vue de la lettre intitulée « Vendredi 17 juillet 1891[7] » et, en dessous, la figure 1b, un extrait du texte brut du document contenant les balises de mise en forme wikicode.

**Figure 1a. Vue partielle de la lettre « Vendredi 17 juillet 1891 » : rendu HTML sur navigateur**

Rechercher sur Une correspondance familiale

Page | Discussion | Lire | Voir le texte source | Voir l'historique

Documents
Lettres
Papiers familiaux et documents divers

Index
Signataires
Destinataires
Personnes citées
Lieux

Ressources
Biographies
Monographies

Recherche sémantique
Point d'accès

Outils
Pages liées
Suivi des pages liées
Pages spéciales
Version imprimable
Lien permanent
Informations sur la page
Chercher les propriétés
Citer cette page

# Vendredi 17 juillet 1891

‹ Dimanche 5 juillet 1891 | Mercredi 19 août 1891 ›

Lettre d'Alphonse Milne-Edwards (Paris) à sa nièce Marie Mertzdorff, épouse de Marcel de Fréville (en villégiature en Bretagne)

*Lettres de tante et oncle Alphonse Milne-Edwards à Mamani*[1]

17 Juillet 1891[2]

Chère Marie

Je reçois ta gentille lettre qui vient me tenir compagnie et en même temps raviver mes regrets de ne pas être avec vous. Le mois de Juillet s'écoule tristement, car le temps a beau couler je retrouve toujours dans l'air, dans la lumière, dans les bruits du dehors des souvenirs pénibles[3]. En ce moment je suis seul à Paris et j'ai trop d'occupations pour m'échapper, c'est à peine si j'ai pu aller le 14 à Launay. Madame Dumas[4] est à Bracancourt où elle restera probablement jusqu'à Lundi et je suis sûr qu'elle aurait prolongé son séjour si elle n'avait pas été trop pressée de reprendre ici son rôle de Maîtresse de Maison ; je regrette qu'elle ne puisse profiter plus longtemps de sa visite à S^aint^-Bon.

Les nouvelles que tu me donnes des enfants[5] me font grand plaisir. Mimi devient forte et sage, elle n'a pas peur de l'eau, elle est raisonnable, elle oublie peut-être aussi de taquiner frères et sœur, c'est bien beau et je te charge de lui faire tous mes compliments. Si elle continue dans cette voie je l'inviterai tous les Mercredis à venir dîner au Jardin des Plantes.

C'est donc le 29 Juillet le grand jour, le jour des émotions[6]. Je pense que Marcel ira surveiller la vente de Livet. Pourvu que les articles des journaux n'aient pas fait venir l'eau à la bouche de trop d'acquéreurs. Quand tu seras fixée sur ce que tu dois faire de ton mois d'Août tu me ~~fée~~r le diras, j'espère toujours que j'en aurai un petit morceau.

Je t'embrasse chère fille ainsi que tous les tiens.

AME

Image produite par les auteurs

**Figure 1b. Vue partielle de la lettre « Vendredi 17 juillet 1891 » : texte brut avec balises wikicode**

```
{{Lettre/Menu_Précédente_Suivante}}
__NOTOC__

<span class="letter-subtitle">Lettre d'Alphonse Milne-Edwards (Paris) à sa nièce Marie Mertzdorff, épouse de Marcel de Fréville
(en villégiature en Bretagne)</span>


[[Fichier:Fs1891-07-17 pages1-4 Alphonse.jpg|150x150px]]    [[Fichier:Fs1891-07-17 pages2-3 Alphonse.jpg|150x150px]]
-----
''Lettres de tante et oncle Alphonse Milne-Edwards à Maman''<ref>Mention ultérieure de la main d'un enfant de Marie
[[Mertzdorff, Charles (1818-1883) et sa famille|Mertzdorff]]-de [[Fréville, Marcel de (1851-1912) et sa famille|Fréville]].
</ref>

17 Juillet 1891<ref>Lettre sur papier deuil.</ref>

Chère Marie

Je reçois ta gentille lettre qui vient me tenir compagnie et en même temps raviver mes regrets de  ne pas être avec vous. Le
mois de Juillet s'écoule tristement, car le temps a beau couler je retrouve toujours dans l'air, dans la lumière, dans les
bruits du dehors des souvenirs pénibles<ref>Aglaé [[Desnoyers, Jules (1800-1887) et ses descendants|Desnoyers]], épouse
d'Alphonse [[Milne-Edwards, Alphonse (1835-1900), son père Henri, leurs proches et descendants|Milne-Edwards]], est décédée le
10 juillet 1887.</ref>. En ce moment je suis seul à Paris et j'ai trop d'occupations pour m'échapper, c'est à peine si j'ai pu
aller le 14 à [[Nogent-le-Rotrou et Launay (L'Aunay d'en-Bas)|Launay]]. Madame [[Dumas, Ernest Charles Jean Baptiste
(1827-1890) et ses descendants|Dumas]]<ref>Cécile [[Milne-Edwards, Alphonse (1835-1900), son père Henri, leurs proches et
descendants|Milne-Edwards]], veuve d'Ernest Charles Jean Baptiste [[Dumas, Ernest Charles Jean Baptiste (1827-1890) et ses
descendants|Dumas]].</ref> est à [[Blaise-Bracancourt (Haute-Marne)|Bracancourt]] où elle restera probablement jusqu'à Lundi et
je suis sûr qu'elle aurait prolongé son séjour si elle n'avait pas été trop pressée de reprendre ici son rôle de Maîtresse de
Maison ; je regrette qu'elle ne puisse profiter plus longtemps de sa visite à [[Ecole Pratique d'Agriculture de Saint-
Bon|S<sup>aint</sup>-Bon]].

Les nouvelles que tu me donnes des enfants<ref>Jeanne, Robert, Charles et Marie Thérèse de [[Fréville, Marcel de (1851-1912) et
sa famille|Fréville]].</ref> me font grand plaisir. Mimi devient forte et sage, elle n'a pas peur de l'eau, elle est
raisonnable, elle oublie peut-être aussi de taquiner frères et sœur, c'est bien beau et je te charge de lui faire tous mes
compliments. Si elle continue dans cette voie je l'inviterai tous les Mercredis à venir dîner au [[Muséum, Jardin des Plantes,
Jardin du roi, Cabinet d'Histoire naturelle du roi|Jardin des Plantes]].

C'est donc le 29 Juillet le grand jour, le jour des émotions<ref>Les de [[Fréville, Marcel de (1851-1912) et sa
famille|Fréville]] s'apprêtent à acheter le [[Château de Livet (Orne)|château de Livet]].</ref>. Je pense que Marcel ira
surveiller la vente de [[Château de Livet (Orne)|Livet]]. Pourvu que les articles des journaux n'aient pas fait venir l'eau à
la bouche de trop d'acquéreurs. Quand tu seras fixée sur ce que tu dois faire de ton mois d'Août tu me <s>l'écr</s> le diras,
j'espère toujours que j'en aurai un petit morceau.

Je t'embrasse chère fille ainsi que tous les tiens.

AME
```

# Exploration : indexation et recherche

24 La navigation dans le corpus est guidée par deux groupes d'index, tous deux accessibles à tout moment depuis un menu flottant. Le premier groupe propose une entrée par catégorie de document : lettres et papiers familiaux, classés par ordre chronologique, et biographies et monographies, triées alphabétiquement.

25 L'intérêt premier qui a dirigé l'étude de la correspondance en tant qu'instrument de sociabilité a conduit à pointer systématiquement les individus, même évoqués évasivement. En résulte un second groupe composé de trois index permettant de naviguer dans l'ensemble documentaire à partir des personnages auteurs (150 noms) ou destinataires (90 noms) de lettres, mais aussi de tous ceux simplement cités explicitement ou de manière allusive (plus de 5 000 noms). Auteurs et destinataires sont indexés dès lors que leur contribution

ne se limite pas à l'ajout d'une très courte formule finale. L'index des personnes citées se veut le plus possible individualisé ; par exemple, lorsqu'un couple est mentionné, l'homme et la femme sont indexés séparément ; lorsqu'il s'agit d'une épouse, la forme « nom de naissance – prénom – nom d'épouse » est utilisée. Cependant, le terme générique « X (famille) » est employé lorsqu'il n'est pas possible de repérer chacun des individus. À chaque individu ainsi pointé correspond une page d'index personnel, recensant les documents édités le mentionnant (voir figure 2).

26 Enfin, un index des lieux complète le tout. Il se limite aux lieux de départ et aux lieux d'arrivée des lettres – que ces lieux soient inscrits ou non par l'auteur. L'index des lieux comporte 160 noms.

27 Le wiki propose par ailleurs une barre de recherche permettant, soit de rechercher dans les titres de pages, soit d'effectuer une recherche plein texte.

## Téléchargement et export

28 La fonctionnalité de téléchargement des pages est déléguée au navigateur Web. Ainsi, il est possible d'exporter une page au format PDF en utilisant l'impression vers PDF de son navigateur. Les fac-similés peuvent être téléchargés depuis la visionneuse d'image accessible en cliquant sur les images affichées en en-tête de chaque lettre.

29 Le lecteur qui souhaite exporter une page du wiki pour l'exploiter dans un autre wiki peut également utiliser la page Special:Export [8] pour obtenir la page au format XML, importable dans un autre wiki depuis la page Special:Import.

## Une API pour interroger et explorer le wiki

30 Le wiki expose à l'adresse lettresfamiliales.ehess.fr/api [9] une interface programmatique documentée pour exécuter des requêtes d'interrogation du contenu, récupérer les métadonnées sémantiques ou obtenir le wikicode ou le rendu HTML des pages. La page spéciale [Spécial:ApiSandbox] [10] propose un espace « bac à sable »

permettant d'expérimenter les différentes actions offertes par l'API du wiki.

## Un corpus structuré par un graphe de connaissances exposé sur le Web des données

31 Le wiki utilise l'extension Semantic MediaWiki qui permet d'enrichir les pages de métadonnées représentées à l'aide du modèle de graphe RDF. En plus d'un espace de rédaction et de publication d'un corpus épistolaire, *lettresfamiliales.ehess.fr* (https://lettresfamiliales.ehess.fr) expose également sur le Web des données un ensemble de métadonnées sous forme de triplets RDF qui modélisent et représentent le graphe social des acteurs du corpus.

32 Chaque page définit ainsi les valeurs d'un ensemble de propriétés issues de vocabulaires RDF standards ou largement utilisés, qui permettent d'ajouter des informations relatives au document et aux individus qui s'y trouvent cités, ou pour les lettres qui en sont auteur ou destinataires. Pour cela, 7 propriétés du vocabulaire de catalogage Resource Description and Access Unconstrained (RDAU) et de l'ontologie bibliographique Bibliographic Ontology (BIBO) ont été choisies.

33 Le tableau 1 donne la liste des propriétés utilisées dans chaque catégorie de document, leurs définitions dans leurs vocabulaires et le sens particulier qui leur est donné dans le contexte du corpus des correspondances familiales.

34 Les métadonnées associées à une page sont affichées en bas de celle-ci dans un panneau dédié (figure 3). En plus de cette présentation, le graphe RDF d'une page est accessible au format RDF/XML à l'aide de la page Spécial:Export_RDF [11].

| Propriété | Définition (rdfs:label) | Nom sur le wiki | Catégories utilisant cette propriété |
|---|---|---|---|
| rdau:P60369 (http://rdaregistry.info/Elements/u/P60369) | *has label* | Titre du document | Toutes |
| rdau:P60493 (http://rdaregistry.info/Elements/u/P60493) | *has other title information* | Sous-titre d'une lettre | Lettre |
| rdau:P60434 (http://rdaregistry.info/Elements/u/P60434) | *has author* | Signataire | Lettre |

| | | | |
|---|---|---|---|
| bibo:recipient (http://purl.org/ontology/bibo/recipient) | *recipient* | Destinataire | Lettre |
| rdau:P60071 (http://rdaregistry.info/Elements/u/P60071) | *has date of production* | Date de rédaction | Lettre |
| rdau:P61039 (http://rdaregistry.info/Elements/u/P61014)[a] | *has related agent of entity* | Individu mentionné | Toutes |
| rdau:P61014 (http://rdaregistry.info/Elements/u/P61014)[b] | *has related place of resource* | Lieu d'expédition ou de réception ; lieu cité | Toutes |

**a**. Cette propriété est nommée mention_agent dans l'export RDF.
**b**. Cette propriété est nommée mention_place dans l'export RDF.

**Figure 3. Affichage des métadonnées définies pour la page de la lettre intitulée « Mercredi 30 août 1820 »**

35 Les métadonnées ajoutées aux documents du wiki permettent de les manipuler à l'aide de requêtes sémantiques pour notamment naviguer dans le graphe des individus du corpus et manipuler les documents à la manière d'une base de données.

36 Plusieurs points d'entrée sont proposés. Avec la page [Spécial:Parcourir][14], il est possible de rechercher la valeur d'une métadonnée pour obtenir l'ensemble des documents la déclarant. La page [Spécial:Requêter][15] permet de construire des requêtes sémantiques complexes à l'aide d'une interface graphique. Enfin, l'API fournie donne également la possibilité d'exécuter des requêtes utilisant la syntaxe de Semantic Mediawiki à l'aide de l'action ask[16].

# Un corpus épistolaire numérique ouvert et librement réutilisable

37 Au-delà des objectifs d'alimentation et de structuration itérative des documents, le projet vise aussi à ouvrir le corpus à la communauté scientifique sous une forme qui maximise son potentiel de réutilisabilité. Cela implique de mettre à disposition le corpus en un ensemble cohérent composé de documents fournis dans des formats interopérables et les plus répandus possible. Or, si MediaWiki s'est montré à l'usage particulièrement adapté pour mener parallèlement l'alimentation en contenu et la structuration du corpus, les fonctionnalités d'export qu'il propose sont surprenamment limitées et peu adaptées à un public non informaticien. L'API est complète et les métadonnées des documents récupérables en RDF par un clic, mais l'interface graphique du wiki ne propose aucun mécanisme permettant de télécharger un sous-ensemble des textes, des fac-similés ou des index. Pour pallier ce manque, un export complet du contenu du wiki est proposé vers plusieurs formats usuels, déposé dans deux entrepôts ouverts de données de la recherche. Ces dépôts sont en outre publiés sous licence libre.

# Du wiki au dépôt de données versionné : export multi-format en masse

38 Les textes structurés en wikicode ont été exportés en HTML à l'aide de l'API du site, puis convertis en document XML standardisés en TEI.

39 La mise en forme des textes dans le wiki reproduisant le plus fidèlement les lettres originales, la typographie et l'organisation des sections et paragraphes, appels de note et notes critiques sont intégralement traduites dans des balises XML.

40 En sus d'une sérialisation XML et HTML, les textes sont également proposés en texte brut, en LaTeX et en PDF. Les textes sont de plus accompagnés des fac-similés au format JPG, dans leur résolution d'origine.

41 Enfin, sont proposés au format CSV les index des lettres, biographies, monographies et papiers familiaux. Chaque ligne des CSV correspond à une page du wiki et contient les métadonnées sémantiques associées. Les intitulés des colonnes de métadonnées reprennent les labels français des propriétés spécialisées pour le wiki et non les noms des prédicats. Le lecteur pourra se rapporter au tableau 1 pour retrouver la correspondance des noms de propriété avec les prédicats RDF.

42 Les index CSV sont placés à la racine du dépôt, tandis que des sous-dossiers nommés d'après les titres des pages contiennent les images et les textes aux différents formats.

43 L'export complet est déposé sur Zenodo (10.5281/zenodo.5810596 (https://doi.org/10.5281/zenodo.5810595)), entrepôt européen de données ouvertes de la recherche [17], et sur l'entrepôt institutionnel de l'École des hautes études en sciences sociales, Didomena (10.25721/wr4c-3e07 (https://doi.org/doi:10.25721/wr4c-3e07)).

# Licence et crédit

44 Le contenu de *Lettresfamiliales.ehess.fr* (https://lettresfamiliales.ehess.fr) et les jeux de données déposés sur Zenodo et Didomena sont publiés sous licence ouverte Etalab 2.0, équivalente à la licence Creative Commons minimale (CC BY [18]). Celle-ci est dite permissive car elle impose pour seule contrainte au réutilisateur de ne pas dévoyer les données et de préciser la date de leur dernière mise à jour [19].

45 Les transcriptions des lettres ayant été produites au sein d'une unité mixte de recherche, placée sous la tutelle de deux établissements publics, et les propriétaires des lettres ayant tous donné leur accord à leur publication, leur réutilisation est libre (Robin 2022), aux conditions de la licence Etalab [20].

46 Le mode de citation des documents – ou du corpus entier – est indiqué dans les mentions légales du wiki [21].

47 Les fac-similés sont pour la majorité issus de numérisations réalisées par les responsables scientifiques du corpus et sont disponibles sous licence ouverte Etalab 2.0. Toutefois, 29 documents iconographiques

ont été fournis par le Muséum national d'histoire naturelle et n'ont pas de licence attribuée.

## Évolutions envisagées

48 Le corpus édité est un objet en construction, régulièrement enrichi de documents, dont certains en cours de publication. Ils complètent le corpus dans sa continuité, de la Révolution française à 1919. De nouvelles biographies, des corrections et des compléments aux textes des lettres antérieures continuent d'être apportés : le site est un chantier qui, à la fois, ajoute des documents nouveaux et révise les anciens.

49 En même temps que la reprise présentée ici, une collaboration entre plusieurs projets d'édition de corpus d'auteurs a donné naissance au groupe de travail « (Ré)utilisabilité » au sein du consortium CAHIER de l'infrastructure de recherche Huma-Num. Le travail dont nous rendons compte ici a alimenté les échanges et réflexions du groupe, qui a récemment publié un vade-mecum rassemblant des recommandations pour favoriser la réutilisabilité des données textuelles dans le cadre des projets de recherche (Aucagne et al. 2022). Cette collaboration a permis d'élargir l'horizon des expérimentations et évolutions à considérer, notamment la création de graphes de réseaux familiaux, de frises historiques dynamiques en lien avec les aspects économiques, politiques et sociaux de l'époque. Nous espérons avoir rendu le corpus épistolaire effectivement et concrètement réutilisable par les chercheurs et ingénieurs en sciences humaines et sociales et nous souhaitons explorer ces nouvelles modalités d'exploration en exploitant le graphe de connaissances construit avec MediaWiki.

50 Il est également prévu que le dépôt complet du corpus sur Zenodo et Didomena soit mis à jour à partir du wiki à un rythme annuel et ponctuellement en cas de modifications importantes.

# Les leçons et attendus d'une expérience de bricolage numérique

51 La reprise intégrale du corpus épistolaire de la famille Mertzdorff-Duméril-Froissart a rapidement et largement dépassé les enjeux techniques liés aux changements de formats et au transfert de plateforme logicielle. Le choix de MediaWiki semble rare pour éditorialiser un corpus de correspondance dans un paysage disciplinaire francophone où se côtoient Omeka, TEI-Publisher, Lodel et de multiples solutions *ad hoc*. Celui-ci était guidé par deux contraintes : d'une part, ne pas bloquer le travail d'alimentation du corpus le temps de la reprise, d'autre part, structurer le corpus existant pour le rendre aisément explorable et exploitable. En plaçant la modélisation des données directement dans les pages elles-mêmes selon une approche déclarative grâce au wikicode, MediaWiki a permis de coupler fortement le travail des ingénieurs et des historiens et de faciliter considérablement les échanges. Ce « bricolage numérique » collectif (Boulaire et Carabelli 2018) a poussé les historiens à « mettre la main à la pâte » en écrivant des instructions en wikicode, et les ingénieurs à s'approprier le corpus. Le résultat premier de ce mouvement d'aller-retour particulièrement fructueux est d'avoir transformé la nature même du corpus, faisant d'une collection de documents un graphe de connaissances explicitement représenté et ouvert sur le Web des données.

52 Les lettres, au carrefour de l'individuel et du social, irriguent tous les domaines, matériels et intimes, politiques, religieux, économiques ou scientifiques ; polyvalentes, elles sont aussi bien un espace d'expression des sentiments qu'un moyen de gestion du quotidien, une arme de séduction ou un instrument de pouvoir. En l'état, et autant que l'on puisse en juger, les lettres et leur appareil critique sont utilisés par au moins deux types de public.

53 D'une part, des lecteurs qui, s'intéressant à une personne ou à une famille (souvent la leur), font des recherches généalogiques. Ces « visiteurs » trouvent dans l'index des personnes citées des noms qui

retiennent leur attention ; ils pointent alors des informations ponctuelles dans les biographies, les monographies et les lettres.

54 D'autre part, les chercheurs qui, travaillant sur un sujet, puisent dans les lettres des éléments qui nourrissent leur recherche. Le contenu informatif varié appelle de nombreux axes de lecture, dans des perspectives historique, anthropologique ou littéraire. Citons deux exemples : l'historienne Catherine Rollet a trouvé dans les lettres familiales, qui parlent souvent des enfants, des traces de pratiques associant parents et enfants (Rollet et Morel 2000) ; en historienne de la littérature, Éléonore Reverzy (2021) s'est intéressée au genre textuel du témoignage, à sa forme narrative, pour la période de la Commune.

55 L'écriture des lettres, acte individuel, tisse des liens entre des personnes et constitue des réseaux au sein de collectivités. MediaWiki, par sa capacité à densément lier les documents entre eux et avec des ressources extérieures, devrait permettre de saisir ces combinaisons réticulaires. On peut, à terme, imaginer la connexion de sites électroniques qui permettrait de multiplier les perspectives, les modes de lecture et les publics.

# Résumé des jeux de données déposés

56 La totalité du corpus édité et publié en janvier 2022 est disponible sous licence Etalab 2.0 sur les entrepôts de données Didomena et Zenodo.

| Entrepôt | Identifiant unique | Intitulé du jeu de données | Date de dépôt |
|---|---|---|---|
| Didomena | doi:10.25721/wr4c-3e07 (https://doi.org/doi:10.25721/wr4c-3e07) | Export multi-format du corpus « S'écrire au XIXe siècle. Une correspondance familiale » édité et publié sur *lettresfamiliales.ehess.fr*. | 04/01/2022 |
| Zenodo | doi:10.5281/zenodo.5810596 (https://doi.org/10.5281/zenodo.5810595) | S'écrire au XIXe siècle. Une correspondance familiale | 04/01/2022 |

57 Le dépôt contient les textes de 3 248 lettres transcrites en autant de fichiers indépendants (aux formats XML, HTML, TXT, LaTeX et PDF),

rassemblés et organisés dans une structure de dossiers par année et décennie. Les quelque 1 002 biographies, 110 monographies et 53 papiers familiaux divers qui composent l'appareil critique sont stockés de la même manière mais organisés par ordre alphabétique. Les facs-similés sont stockés dans le même dossier que le texte de la lettre. Le tout est archivé au format ZIP dans le fichier data.zip.

58 Les index catégoriels qui structurent le wiki sont exportés en autant de fichiers CSV, lesquels contiennent également les métadonnées associées à chaque document.

59 Les données sont disponibles sous licence Etalab 2.0, à l'exception de quelques images fournies par le Muséum national d'histoire naturelle, dont la liste est disponible dans le fichier README.md.

60 Aucagne, Julie, Marguerite Bordry, Camille Desiles, Francine Filoche, Anne Garcia-Fernandez, Greslou Elisabeth, Camille Koskas, Gwenaëlle Patat, Richard Walter, et Pierre Willaime. 2022. « Vade-mecum pour la réutilisabilité des données ». Paris: IR Huma-Num, Consortium Cahier, Groupe de travail Réutilisabilité. https://hal.archives-ouvertes.fr/hal-03698675.

61 Boulaire, Cécile, et Romeo Carabelli. 2018. « Du *digital naive* au bricoleur numérique : les images et le logiciel Omeka ». In *Expérimenter les humanités numériques. Des outils individuels aux projets collectifs*, édité par Étienne Cavalié, Frédéric Clavert, Olivier Legendre, et Dana Martin, 81-103. Montréal: Presses de l'université de Montréal. https://doi.org/10.4000/books.pum.11115.

62 Bourdelais, Patrice. 1995. « Les banques de données au CRH – Présentation ». *Les Cahiers du Centre de recherches historiques* 14-15. https://doi.org/10.4000/ccrh.2657.

63 Dauphin, Cécile, Pierrette Lebrun-Pézerat, et Danièle Poublan. 1991. « L'enquête postale de 1847 ». In *La Correspondance. Les usages de la lettre au* XIX*e siècle*, édité par Roger Chartier, 11-119. Paris: Fayard.

64 Dauphin, Cécile, Pierrette Lebrun-Pézerat, et Danièle Poublan. 1995. *Ces Bonnes Lettres. Une correspondance familiale au* XIX*e siècle*. Paris: Albin Michel.

65 Furet, François, et Jacques Ozouf. 1977. *Lire et écrire. L'alphabétisation des français de Calvin à Jules Ferry*. Paris: Éditions de Minuit.

66 Lévy-Dumoulin, Olivier. 2005. « L'enquête collective en lieu d'identité ». *Les Cahiers du Centre de recherches historiques* 36. https://doi.org/10.4000/ccrh.3062.

67 Reverzy, Éléonore. 2021. *Témoigner pour Paris. Récits du Siège et de la Commune (1870-1871). Anthologie*. Paris: Kimé.

68 Robin, Agnès. 2022. *Droit des données de la recherche. Science ouverte, innovation, données publiques*. Louvain-la-Neuve: Larcier.

69 Rollet, Catherine, et Marie-France Morel. 2000. *Des bébés et des hommes. Traditions et modernité des soins aux tout-petits*. Paris: Albin Michel.

70 Triaire, Paul. 1892. *Bretonneau et ses correspondants. Ouvrage comprenant la correspondance de Trousseau et de Velpeau avec Bretonneau*. Paris: Alcan.

## NOTES

1 Charles Mertzdorff, un industriel alsacien (1818-1883), Paris, chez l'auteur, 1983 ; André Constant Duméril, médecin et naturaliste (1774-1860), Paris, chez l'auteur, 1984 ; Histoire d'une vieille maison de famille à Campagne-lès-Hesdin, Paris, chez l'auteur, 1985.

2 À l'exception des sauts de ligne et de l'écriture en marge, repérables sur les fac-similés.

3 Les formes les plus contemporaines ont ainsi été privilégiées : tems a été transformé en temps, etc.

4 https://correspondancefamiliale.ehess.fr.

5 Content Management System. Les plus connus sont Wordpress et Drupal

6 Resource Description Framework. RDF est le modèle de représentation standard des données liées sur le Web de données.

7 http://lettresfamiliales.ehess.fr/view/Vendredi_17_juillet_1891

8 Les pages de l'espace de nom « Spécial » sont des pages utilitaires permettant d'effectuer une action sur la page dont le titre est donné à la suite. Par exemple, https://lettresfamiliales.ehess.fr/view/Spécial:Export/Vendredi_12_novembre_1784 permet d'exporter la lettre intitulée « Vendredi 12 novembre 1784 » au format XML.

9 https://lettresfamiliales.ehess.fr/api.php

10 https://lettresfamiliales.ehess.fr/view/Spécial:ApiSandbox

11 Par exemple, https://lettresfamiliales.ehess.fr/view/Spécial:Export_RDF/Vendredi_12_novembre_1784.

14 https://lettresfamiliales.ehess.fr/view/Spécial:Parcourir

15 https://lettresfamiliales.ehess.fr/view/Spécial:Requêter

16 Voir la page d'aide sur le site de Semantic MediaWiki (https://www.semantic-mediawiki.org/wiki/Help:API:ask). La documentation du langage de requêtes de Semantic MediaWiki est disponible à l'URL https://www.semantic-mediawiki.org/wiki/Help:Inline_queries.

17 Zenodo a été créé et est maintenu par les chercheurs du Conseil européen pour la recherche nucléaire (CERN), qui regroupe la quasi-totalité des pays d'Europe de l'Ouest, membres ou non de l'Union européenne.

18 Etalab est l'une des deux seules licences de réutilisation pouvant être choisies pour des données issues d'une recherche menée avec au moins 50 % de fonds publics : cf. art. D323-2-1 du Code des relations entre le public et l'administration (CRPA) (https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000034504991/). Une liste spécifique existe pour les licences attribuables aux logiciels.

19 Art. L322-1 du CRPA (https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000032255220/).

20 « Dès lors que les données issues d'une activité de recherche financée au moins pour moitié par des dotations de l'État […] ne sont pas protégées par un droit spécifique ou une réglementation particulière et qu'elles ont été rendues publiques par le chercheur, l'établissement ou l'organisme de recherche, leur réutilisation est libre » : loi « Pour une République numérique », art. 11 (https://www.legifrance.gouv.fr/jorf/article_jo/JORFARTI000033203004/) (art. L544-4 du Code de la recherche (https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033205794/)). Cette disposition vient en complément de l'article L321-3 du CRPA (https://www.legifrance.gouv.fr/codes/article_lc/LEGIARTI000033205561/).

21 https://lettresfamiliales.ehess.fr/view/Une_correspondance_familiale:Mentions_légales.

# RÉSUMÉS


## Français

Le site *lettresfamiliales.ehess.fr* (https://lettresfamiliales.ehess.fr/) est un wiki sémantique qui présente une version actualisée et augmentée de la correspondance écrite de la famille Mertzdorff-Duméril-Froissart. Cette famille regroupe des industriels du textile, des scientifiques du Muséum national d'histoire naturelle de Paris et des hauts fonctionnaires civils et militaires qui réalisent une ascension sociale entre la fin du XVIIIe siècle et le début du XXe siècle. Leur correspondance est un véritable instrument de sociabilité, un moyen d'accroître leur statut social et de conserver la trace de cette élévation dans la société française du XIXe siècle. L'édition numérique de plus de 3 000 lettres entièrement transcrites, initialement publiées sur le site Web *correspondancefamiliale.ehess.fr*, était menacée à court terme en raison de l'obsolescence de son logiciel support Lodel. L'aboutissement de ce travail est une publication renouvelée, utilisant Semantic MediaWiki, qui expose ce corpus dans une version enrichie d'un grand nombre de relations sémantiques liant les lettres et leurs acteurs au sein d'un graphe de connaissances (*knowledge graph*) exposé sur le Web des données. Ces liens sont exploités pour construire une base de données des personnes mentionnées dans les lettres, des auteurs et des lieux d'expédition, mise à jour dynamiquement au fil de l'enrichissement du corpus. L'ensemble des textes et des fac-similés numérisés est disponible sous licence ouverte sur les entrepôts de données de la recherche Zenodo (doi : 10.5281/zenodo.5810596 (https://doi.org/10.5281/zenodo.5810595)) et Didomena (doi : 10.25721/wr4c-3e07 (https://doi.org/doi:10.25721/wr4c-3e07)).

## English

This article reports on the process of recovery, restructuring and digital publication of the written correspondence of the Mertzdorff-Duméril-Froissart family in the form of a semantic wiki available at *lettresfamiliales.ehess.fr* (https://lettresfamiliales.ehess.fr/). Members of this family include textile manufacturers, scientists from the National Museum of Natural History and high rank civil and military officials who achieved a social ascension between the end of the 18th century and the beginning of the 20th century. Their correspondence is a powerful tool of sociability, a means to increase their social status and to keep track of their rise in the Parisian society of the 19th century. The digital edition of these more than 3000 fully transcribed letters, initially published on the website *correspondancefamiliale.ehess.fr* (http://correspondancefamiliale.ehess.fr/), is threatened in the short term due to the end of life of its support software Lodel. The outcome of this work is a renewed publication following FAIR principles, using Semantic MediaWiki, exposing this corpus in a version enriched with a large number of semantic relations linking the letters and their actors within a knowledge graph exposed on the Web of data. These

links are exploited to build a database of authors and recipients along with people and places mentioned in the letters. The digitised texts and facsimiles are available under the French « Open Licence » on the repositories Zenodo (doi: 10.5281/zenodo.5810596 (https://doi.org/10.5281/zenodo.5810595)) and Didomena (doi: 10.25721/wr4c-3e07 (https://doi.org/doi:10.25721/wr4c-3e07)).

## INDEX

**Mots-clés**
corpus de lettres, édition critique, données ouvertes, Web sémantique

**Keywords**
corpus of letters, critical edition, open data, Semantic Web

## AUTEURS

**Bertrand Duménieu**
UMR 8558 Centre de recherches historiques, EHESS, Paris, France
Bertrand Duménieu est ingénieur de recherche en sciences de l'information géographique. Ses activités d'ingénierie et de recherche portent sur l'extraction, la spatialisation et l'analyse de données géographiques historiques. Il participe également à plusieurs travaux de publication numérique de corpus historiques au sein du Centre de recherches historiques.
ORCID 0000-0002-2517-2058 (https://orcid.org/0000-0002-2517-2058)

**Danièle Poublan**
UMR 8558 Centre de recherches historiques, EHESS, Paris, France
Danièle Poublan, ingénieure d'études, a travaillé sur les écritures ordinaires, les images de grande circulation et les relations entre femmes et hommes dans la France du xixe siècle.
ORCID 0000-0001-5662-2376

**Jean-Damien Généro**
UMR 8558 Centre de recherches historiques, CNRS, Paris, France
Jean-Damien Généro est ingénieur d'études, en charge de la gestion des projets numériques du Centre de recherches historiques.
ORCID 0000-0002-0754-9130 (https://orcid.org/0000-0002-0754-9130)

**Francine Filoche**
UMR 8558 Centre de recherches historiques et Pôle numérique recherche de la direction des systèmes d'information, EHESS, Paris, France
Francine Filoche est chargée de gestion de données et d'accompagnement de projets de recherche. Bibliothécaire, son périmètre d'action s'est élargi à l'assistance à la gestion de données notamment à la rédaction de plans de gestion

de données de projet ou de structure. Elle facilite la prise en main d'outils de signalement ou de plateformes de dépôt de productions scientifiques et sensibilise au respect des bonnes pratiques en faveur de la science ouverte.
ORCID 0000-0001-6170-9820 (https://orcid.org/0000-0001-6170-9820)

**Patricia Bleton**
UMR 8558 Centre de recherches historiques, EHESS, Paris, France
Patricia Bleton est chargée de ressources documentaires au Centre de recherches historiques de l'EHESS.