Du livre imprimé au Web sémantique : le projet du *Dictionnaire des éditeurs français du XIX^e siècle*

From the Printed Book to the Semantic Web: Towards a Dictionary of 19th-century French Publishers

Jean-Charles Geslot et Viera Rebolledo-Dhuin

Édition électronique

URL : <https://journals.openedition.org/revuehn/426>

DOI : 10.4000/revuehn.426

ISSN : 2736-2337

Éditeur

Humanistica

Référence électronique

Jean-Charles Geslot et Viera Rebolledo-Dhuin, « Du livre imprimé au Web sémantique : le projet du *Dictionnaire des éditeurs français du XIX^e siècle* », *Humanités numériques* [En ligne], 2 | 2020, mis en ligne le 01 juin 2020, consulté le 15 juillet 2021. URL : <http://journals.openedition.org/revuehn/426> ; DOI : <https://doi.org/10.4000/revuehn.426>

Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.

Du livre imprimé au Web sémantique : le projet du *Dictionnaire des éditeurs français du XIX^e siècle*

From the Printed Book to the Semantic Web: Towards a Dictionary of 19th-century French Publishers

Jean-Charles Geslot et Viera Rebolledo-Dhuin

Résumés

Cet article vise à présenter les enjeux techniques du projet *DEF19* (*Dictionnaire des éditeurs français du XIX^e siècle*). Il montre comment les choix scientifiques opérés pour définir le périmètre de ce dictionnaire (définition de ce qu'est un éditeur, recensement de celles et ceux qui entrent dans cette catégorie...) ont en grande partie déterminé les solutions techniques. Il décrit les diverses expériences menées pour mettre au point la base de données, à savoir les logiciels expérimentés et les problèmes qui ont été rencontrés (intégration de données hétérogènes issues de bases différentes, notamment), ainsi que la façon dont ces derniers ont pu être résolus. Il analyse l'intérêt que le Web sémantique, à travers l'usage de la plateforme Omeka S, peut représenter pour une telle entreprise scientifique, en montrant également les diverses potentialités qui se sont ouvertes au fil de la réalisation du projet en termes de collaborations techniques et institutionnelles.

This article aims to describe technical issues of *DEF19* project (*Dictionnaire des éditeurs français du XIX^e siècle*, i.e. *19th-century French Publishers Dictionary*). Scientific choices have been made (what is a publisher and who is a publisher in 19th-century France?), which have widely affected technical answers. Several software solutions have been considered to build the database, each of them posing specific problems (and resulting in different solutions), particularly in integrating data coming from various preexisting databases. Our final choice, relying on Semantic Web

technology through the Omeka S platform, offers many development potentials for such an academic challenge. These possibilities are described by the authors, together with technical and institutional cooperations allowed by this tool and gradually unveiled since the beginning of the project.

Entrées d'index

MOTS-CLÉS : histoire, histoire du livre, base de données, constitution de corpus, dictionnaire, éditeur, prosopographie, Web, Web sémantique

KEYWORDS: history, history of books, database, corpus building, dictionary, publisher, prosopography, Web, semantic Web

Introduction

¹ L'histoire du livre est un domaine particulièrement propice à l'étude de données¹. La tradition de l'analyse statistique, par exemple, y est forte et ancienne : recenser et compter les diverses éditions d'un ouvrage, les exemplaires imprimés ou vendus, les lieux de leur diffusion (librairies, bibliothèques, cabinets de lecture...), les lecteurs eux-mêmes, ont constitué dès longtemps un axe important de recherches, structuré notamment autour de deux perspectives bien identifiées, la bibliométrie (Vaillant 1992) et la prosopographie des gens du livre. Le nombre relativement important de bases de données couvrant aujourd'hui ces deux domaines témoigne de leur vitalité. Du côté du recensement des œuvres, plusieurs projets, passés ou en cours, offrent des ressources nombreuses sur des champs plus ou moins larges de la production imprimée, qu'il s'agisse de genres comme le roman d'anticipation², le récit de voyage³, la littérature de mystères urbains⁴ ou la publicité⁵ ; d'imprimés particuliers comme la presse⁶ ou les catalogues de vente de bibliothèques⁷ ; ou bien encore de la production d'éditeurs comme Mame⁸, Guillaumin⁹, ou à l'échelle d'un État comme le Québec¹⁰.

² Les gens du livre et de l'imprimé, de leur côté, ont depuis plusieurs années fait l'objet d'entreprises de recensement aboutissant à des répertoires et à des dictionnaires ainsi qu'à des analyses prosopographiques. Deux vastes enquêtes collectives font figure d'entreprises pionnières pour le champ français. D'une part, le vaste répertoire des imprimeurs-libraires de 1500 à 1810, lancé dans les années 1980 par la Bibliothèque nationale de France (BNF), sous la direction notamment de Jean-Dominique Mellot, et qui a abouti à quatre publications successives entre 1988 et 2001, offrant un nombre de notices croissant jusqu'à 5 200 dans la dernière version du répertoire. D'autre part, l'enquête sur la librairie du XVIII^e siècle menée au sein de l'Institut d'histoire moderne et contemporaine (IHMC, UMR 8066), et dirigée par Frédéric Barbier et Sabine Juristic. *In fine*, les chercheurs disposent aujourd'hui d'au moins huit dictionnaires et répertoires imprimés recensant les professionnels du livre, soit de façon globale (Bibliothèque nationale de France 2004), soit

sur des espaces restreints comme Paris (Barbier, Juratic, et Mellerio 2007) ou le Nord (Barbier 2002), soit pour certaines catégories particulières – éditeurs de musique (Devriès et Lesure 1979), éditeurs d'estampes de Paris (Préaud *et al.* 1987) et de Lyon (Martin-de-Vesvrotte, Perez et Pommier 2002), femmes libraires (Arbour 2003). Plusieurs bases de données consultables en ligne complètent ces publications, sur des aires géographiques et des catégories professionnelles diverses : entre autres exemples, les notices biographiques des gens du livre à Lyon au XVIII^e siècle dues à Dominique Varry¹¹, le Dictionnaire des imprimeurs-lithographes du XIX^e siècle, dirigé par Corinne Bouquin et Élisabeth Parinet¹², la rubrique « Livre-papier-carton » dans le cadre du *Maitron* en ligne¹³, le Dictionnaire des journalistes francophones du XIX^e siècle de la plateforme Médias 19¹⁴ ; et, pour l'étranger, le Répertoire des imprimeurs et éditeurs suisses actifs avant 1800¹⁵, la base des gens du livre au Québec¹⁶, ou celle établie par Ian Maxted sur les acteurs du commerce du livre en Angleterre¹⁷, sans compter le portail sur les éditeurs ibéro-américains aux XIX^e-XXI^e siècles porté par Pura Fernandez et hébergé par la bibliothèque virtuelle Miguel de Cervantes¹⁸.

FIGURE 1. BASES DE DONNÉES ET DICTIONNAIRES PROSOPOGRAPHIQUES FRANÇAIS EN HISTOIRE DU LIVRE

Jean-Charles Geslot, décembre 2017

3

C'est dans la continuité de cette tradition de bases de données sur les gens du livre que se situe le programme ANR DEF19 « *Dictionnaires des éditeurs français du XIX^e siècle* ». Celui-ci s'est résolument inscrit, dès son lancement, dans les perspectives offertes par les humanités numériques, en prévoyant la constitution d'une base de données de référence. Au fur et à mesure de son avancée cependant, des potentialités nouvelles n'ont cessé de s'ouvrir, tant en termes scientifiques et techniques que de partenariats. Si la démarche a été quelque peu empirique, elle n'en a pas moins permis d'aboutir aujourd'hui à un résultat bien plus ambitieux que ce qui avait pu être imaginé au départ. L'objet de cet article sera donc de décrire comment le projet initial s'est progressivement étoffé pour tenter de mettre au service de la recherche en histoire de l'édition et de ses métiers toutes les potentialités offertes par les outils du Web sémantique¹⁹.

Genèse : du projet à la base de données

Le projet *DEF19* : enjeux institutionnels et scientifiques

4

Le projet de Dictionnaire des éditeurs du XIX^e siècle a pris forme en 2014 au sein du Centre d'Histoire culturelle des sociétés contemporaines (CHCSC, EA 2448). Créé en 1992 à l'université de Versailles-Saint-Quentin-en-Yvelines (UVSQ), ce laboratoire a depuis son ouverture inscrit l'histoire du livre et de l'édition parmi les axes principaux de ses travaux, sous la direction du professeur Jean-Yves Mollier qui, outre ses propres publications, a fait soutenir dans ce domaine de nombreux mémoires et thèses. Parmi celles-ci, il faut tout de suite évoquer celles de Viera Rebolledo-Dhuin (2011) et de Marie-Claire Lefils-Boscq (2013), consacrées au monde des libraires et des métiers du livre parisiens du XIX^e siècle : toutes deux ont édifié dans le cadre de leur travail doctoral des bases de données sur les professionnels du livre qui ont servi de matrice à celle du projet *DEF19*, lequel constitue en quelque sorte le prolongement et l'approfondissement de ces recherches. Toutefois, l'objectif n'était pas seulement de constituer une base sur les éditeurs du XIX^e siècle : il s'agissait aussi de fournir un outil « papier », un dictionnaire correspondant à un résultat plus classique, plus traditionnel, de ce genre d'entreprises, en même temps qu'à un horizon d'attente des publics potentiellement intéressés par un tel outil (bibliothécaires, bibliophiles, historiens du livre et de l'édition, etc.).

5

Le contenu de ce double dictionnaire, qui vise à recenser l'ensemble des individus et des établissements ayant édité au moins un ouvrage entre 1800 et 1914²⁰, pose de nombreuses questions, tant d'un point de vue terminologique que scientifique. Qu'est-ce qu'un éditeur ? Qu'est-ce qu'un livre ? Autant d'interrogations déjà anciennes parmi les spécialistes du livre (Mellot 2006) qu'il ne s'agit pas ici de développer, mais dont l'évocation est nécessaire, tant les choix opérés par l'équipe *DEF19* pour y faire face ont pu déterminer la mise en place des solutions techniques. Comment recenser les éditeurs, sachant que cette profession, même si elle s'impose comme la cheville ouvrière de l'industrie du livre à partir des années 1830, constitue longtemps une catégorie en voie d'objectivation, aux contours encore flous face aux anciennes figures dominantes du monde de l'imprimé que sont les imprimeurs, imprimeurs-libraires et libraires-éditeurs ? Sachant également que la législation, particulièrement normative dans le domaine de l'imprimé depuis le Premier Empire, ne reconnaît pas officiellement la profession d'éditeur ? Le décret de 1810 impose en effet aux professionnels du livre l'obtention d'un brevet pour pouvoir exercer leur activité (Sorel et Leblanc 2008). Ce brevet, délivré par les autorités, est accordé à des individus au titre d'« imprimeur » ou de « libraire²¹ ». Jusqu'à sa suppression en septembre 1870, et même un peu au-delà, ce régime dit « du brevet » impose sa marque normative au monde du livre. L'éditeur, qui n'existe pas pleinement aux yeux des autorités²², est une catégorie transversale, qui concerne autant des imprimeurs que des libraires, ces deux catégories n'étant pas elles-mêmes séparées par des frontières étanches. L'avantage, cependant, réside dans le fait que les historiens possèdent, grâce

aux archives de la direction de l’Imprimerie et de la Librairie au ministère de l’Intérieur (série F18 des Archives nationales), le moyen de recenser de façon exhaustive les imprimeurs et libraires brevetés en France dans les trois premiers quarts du XIX^e siècle, et d’y retrouver ceux de ces professionnels qui correspondent bien à la définition d’éditeurs, c’est-à-dire qui ne se contentent pas, comme les imprimeurs, de fabriquer des livres, ni, comme les libraires, de les vendre, mais se chargent également de les concevoir, de recruter les auteurs, de définir les meilleures stratégies de la diffusion des ouvrages, etc. L’une des premières tâches du projet *DEF19* devait donc consister à établir la liste de ceux qui appartenaient à la catégorie « éditeurs » dans l’ensemble plus vaste que constituent les professionnels du livre : ce qui ne devait pas manquer d’avoir, comme nous le verrons, de nombreuses implications techniques.

6 Le projet *DEF19* a été présenté dans le cadre de l’appel à projets générique 2014 de l’Agence nationale de la recherche relatif aux « grands défis sociétaux » et au « défi de tous les savoirs ». Au terme de l’examen de son dossier, il s’est vu accorder un budget de 200 000 € sur quatre, puis cinq ans (2014-2019). Une équipe d’une dizaine de personnes s’est constituée, rassemblant chercheurs, bibliothécaires et archivistes issus d’institutions différentes : le CHCSC de l’UVSQ, le pôle Métiers du livre de l’université Paris-Nanterre, le service à la Recherche de la bibliothèque universitaire de Saint-Quentin-en-Yvelines, le service de l’Inventaire rétrospectif de la BNF, le département Justice et Intérieur des Archives nationales²³. Lancé en octobre 2014, ce projet devait se terminer en octobre 2019 avec la publication du dictionnaire papier. Celui-ci couvre la période allant de 1800 – date de la réorganisation des métiers du livre par le Consulat (Vouillot 1990) – aux années 1870 – fin du régime du brevet. La base de données, hébergée par la TGIR Huma-Num et destinée à vivre bien au-delà de l’échéance de 2019, porte, elle, sur une période plus large, allant des années 1780 à la Grande Guerre.

7 Afin d’alimenter la réflexion sur le projet scientifique, trois manifestations ont également été programmées. Une première journée d’études s’est déroulée en octobre 2015 à la bibliothèque de l’Arsenal. Intitulée « Dictionnaires et répertoires des gens du livre, en Europe et dans le monde. Expériences et perspectives », elle a permis de confronter les retours d’expérience de chercheurs et de chercheuses impliqué(e)s dans des projets similaires, en France et à l’étranger (Italie, Québec, Royaume-Uni), et d’envisager les outils et les possibilités de telles enquêtes dans d’autres espaces (aire germanique, Australie, Brésil, Espagne, Suisse, Tunisie²⁴). Une deuxième journée s’est tenue à l’université de Saint-Quentin-en-Yvelines en février 2017. Consacrée cette fois aux « Marges et périphéries du monde du livre », elle a tenté de croiser les approches pour appréhender les domaines du monde de l’imprimé moins faciles à étudier, ceux des « petits » éditeurs, des éditeurs provinciaux, des producteurs d’imprimés particuliers (partitions, éphémères, atlas...), et pour appréhender l’activité de l’édition dans ses liens vitaux avec la papeterie et l’imprimerie. Ces deux événements ont ainsi permis de dégager un certain nombre de problématiques, d’un point de vue méthodologique notamment, susceptibles d’aider à la réalisation de la base de données *DEF19*, notamment par les témoignages de chercheurs et de chercheuses engagé(e)s dans des projets de bases de données similaires, comme Josée Vincent pour le Dictionnaire des gens du livre au Québec

ou Marie-Cécile Bouju pour le Maitron des ouvriers du livre et du papier. Un colloque de clôture du projet s'est, enfin, tenu aux Archives nationales les 4 et 5 avril 2019. Intitulé « Être éditeur au XIX^e siècle », il a permis de présenter des travaux effectués dans la lignée des perspectives biographiques et prosopographiques de *DEF19*, et d'exposer une série de premiers résultats issus de l'exploitation de la base de données.

La base *DEF19* : à la recherche de l'outil idéal. Du Web au Web sémantique

⁸ L'approche empirique menée par Viera Rebolledo-Dhuin, post-doctorante chargée de la mise en place de la base de données entre 2015 et 2017, en collaboration avec le service à la Recherche de la bibliothèque universitaire de l'université de Versailles-Saint-Quentin-en-Yvelines (BU-UVSQ), a conduit à tester différents logiciels avant d'adopter Omeka S.

⁹ Le personnel de la BU-UVSQ utilisant déjà le logiciel QES, nous avons réfléchi, dans une première étape, aux possibilités d'utiliser cet outil payant, conçu par la société QWAM, pour la gestion et le partage des flux numériques documentaires²⁵. Malheureusement, les développements de ce logiciel auraient été trop complexes et trop coûteux pour notre entreprise. C'est pourquoi William Demet et Viera Rebolledo-Dhuin ont alors enquêté sur un logiciel libre adapté au projet. Parallèlement, d'autres membres de l'équipe ont mis en place une base de données développée sous MySQL. Cependant, l'hébergement en local de cette base compliquait la saisie collective simultanée et ne permettait pas de faire des liens entre les notices d'éditeurs, ni même de lier les maisons d'édition aux éditeurs y participant.

¹⁰ Suite à ces premiers essais, le service à la recherche de la BU-UVSQ est entré en contact avec la TGIR Huma-Num²⁶, qui a non seulement proposé d'héberger notre base de données sur son serveur, mais également offert la possibilité de tester deux logiciels, plus appropriés à notre objet et à nos sources. Nous avons ainsi successivement pris en main les logiciels Atom²⁷ et Omeka 2.4. Le premier, dont le nom provient de la contraction de *Access To Memory*, présentait plusieurs avantages puisqu'il s'agit d'un logiciel libre reposant sur un code source ouvert – et donc modifiable –, fondé sur un environnement multilingue et à multiples référentiels tout en facilitant la saisie de relations réciproques. Ainsi, les relations hiérarchiques et celles entre pairs, qu'il s'agisse de relations de parenté ou de relations professionnelles, étaient parfaitement aisées à intégrer : « est le fils de » devenait automatiquement « est le père de » ou « est la mère de » dans la fiche associée, tandis que « est le frère/la sœur de » était dupliqué ; de même, « est employé par » devenait « est l'employeur de », tandis que « est l'associé de » se répliquait de la même façon. Néanmoins, Atom, étant conçu par et pour les archivistes et pour la description archivistique – avec le soutien du Conseil international des archives (ICA²⁸) – selon les normes internationales (XML-EAD, XML/EAC-CPF, etc.), le détournement des champs existants en dehors de ces normes internationales s'avérait très complexe et la création de champs – numériques notamment – impossible.

Après plusieurs mois de développement sous Atom, ce qui nous a notamment permis d'entrer en contact avec le service des archives de la Maison archéologie et ethnologie (MAE, université Paris-Nanterre), utilisateur assidu dudit logiciel²⁹ –, nous avons testé Omeka (dans sa version classique, version 2.4³⁰). Ce logiciel libre à code source ouvert, développé par le Roy Rosenzweig Center for History and New Media³¹ – également fondateur du logiciel bibliographique bien connu, Zotero, et de celui de gestion d'images annotées, Tropy –, se situe à l'intersection entre les CMS (logiciels de contenu comme WordPress), la gestion d'archives et celle de bibliothèques, et est particulièrement utilisé, à ce jour, pour les expositions virtuelles. Il laissait donc une grande marge de manœuvre pour le développement de site adapté au projet désiré. Par ailleurs, la communauté montante – notamment francophone, ce qui n'est pas le cas pour Atom – participe à la multiplication de plugins, ce qui laissait beaucoup d'espoirs de développements adaptés à nos besoins. Pour autant, lors de nos manipulations, nous nous sommes heurtés à l'impossibilité de créer des blocs de champs répétables, nécessaires notamment à la prise en compte des brevets issus de la série F18, conservée aux Archives nationales (AN³²). C'est ainsi que nous avons contacté l'entreprise Limonade & Co³³ – laquelle travaille, par ailleurs, avec Huma-Num, notamment pour le pack Nakalona³⁴ –, qui nous a proposé de développer notre base de données sous Omeka S³⁵, la version sémantique du précédent logiciel. Outre les avantages présentés par la version classique d'Omeka, l'adoption de la version sémantique visait à résoudre plusieurs problèmes liés à la précédente version, qui risque aujourd'hui de disparaître. Ainsi, la version sémantique d'Omeka nous a permis de créer des blocs de champs répétables à l'envi (brevets, adresses, etc.), ce qui est indispensable pour la prise en compte du fait que les gens du livre peuvent être brevetés – successivement ou simultanément – en tant que libraire, imprimeur ou lithographe.

TABLEAU 1. LOGICIELS TESTÉS POUR LA MISE EN PLACE DE LA BASE DE DONNÉES DEF19

Période	Logiciel	Hébergement	Atouts	Inconvénients
Étape 1 (2014 à 2015)	QES	Serveur BU-UVSQ	Logiciel utilisé par la BU-UVSQ	Logiciel payant ; développements complexes et coûteux car totalement externalisés
	MySQL	Local	Logiciel libre, développé en interne gratuitement	Hébergement local ne permettant pas la saisie simultanée, absence de lien entre les notices
Étape 2 (2015 à 2016)	Atom	Serveur TGIR Huma-Num	Logiciel libre ; création très aisée de liens entre notices et notamment de liens réciproques (hiérarchiques ou entre pairs)	Logiciel créé par et pour les archivistes conformément aux normes internationales d'archivistique ; détournement des champs existants difficile, impossibilité de créer des champs en dehors de ce cadre.
	Ome-ka 2.4		Logiciel libre ; création aisée de liens entre notices ou de champs de nature variée (liste déroulante, texte, date, etc.) ; nombreux plugins disponibles	Impossible de créer des blocs de champs répétables (ex : brevets)
Choix	Ome-ka S		Logiciel libre ; création de liens entre notices ou de champs de nature variée ; création de blocs de champs répétables ; nombreux plugins disponibles	Version alpha du logiciel au moment du développement ; plugins en cours de développement

Viera Rebolledo-Dhuin

12

Compte tenu du fait qu'Ome-ka S est en cours de développement et que ses utilisateurs sont encore peu nombreux, l'entreprise Limonade & Co a développé notre base de données sous la version alpha et n'a pas pu mettre en place tous les plugins souhaités – notamment ceux pour l'export des données (au format CSV, par exemple) pour leur traitement quantitatif. Cela étant dit, l'adoption de la version sémantique d'Ome-ka offre à notre travail de grandes potentialités, en matière de visibilité mais aussi du point de vue de l'enrichissement des données et des collaborations scientifiques et techniques nécessaires au programme DEF19.

Collecte, structuration et intégration des données

- 13 Parallèlement à la recherche de l'outil adapté à notre projet, Viera Rebolledo-Dhuin s'est également chargée de la collecte, de la structuration et de l'intégration des données existantes tout en cherchant des collaborations extérieures pour enrichir les données du dictionnaire – ou accroître leur visibilité – à plus long terme.

Intégration des données existantes immédiatement disponibles

- 14 Le programme *DEF19* s'appuie en premier lieu sur deux fonds principaux, conservés dans deux des institutions partenaires du projet : les dossiers de brevets conservés dans la série F18 du département Justice et Intérieur des AN et les notices nationales d'autorité rédigées par les membres du département de l'Inventaire rétrospectif de la Bibliothèque nationale de France.
- 15 Au sein des AN, Patrick Laharie (+) et Odile Krakovitch avaient mis en place deux bases de données, qui devaient intégrer la base QUIDAM – librement consultable sur les sites des AN – mais qui, de fait, n'ont jamais été publiées : Libredit, pour les brevetés parisiens, et Gutenberg, pour les brevetés français. Viera Rebolledo-Dhuin, ayant travaillé avec Patrick Laharie et ayant pu récupérer le fichier PDF de Libredit au début de sa thèse³⁶, pour le transformer en tableur et le compléter, cette ressource a pu constituer un des fondements de la base de données *DEF19*. S'y est bientôt ajoutée, la base Gutenberg, livrée peu avant l'été 2015 sous forme de fichier texte, et qui, au moment de la publication de cet article, était en cours de transformation pour l'intégration de données structurées et liées dans la base³⁷.
- 16 Le département de l'Inventaire rétrospectif de la BNF, dirigé par Jean-Dominique Mellot, a par ailleurs transmis, dès le début de notre entreprise, l'ensemble des notices nationales d'autorité des éditeurs du XIX^e siècle alors disponibles, qu'il s'agisse de personnes physiques, morales ou de collectivités. Le département rétrospectif corrigeant, complétant et mettant à jour régulièrement les notices d'autorité, il a fallu réfléchir à ces transmissions successives, sur lesquelles nous reviendrons et que résout en partie l'utilisation du Web sémantique.
- 17 Enfin, lors de cette première étape de travail sur les données, il s'agissait de prendre en compte les données intégrées dans la base provisoire développée sous MySQL par Marie-Claire Lefils-Boscq et Gérard Martinnelli. La première « collecte » s'appuie finalement sur des données bien structurées, mais de manière différente. Il a donc fallu harmoniser cette structuration, avant de procéder à un travail d'identification – les femmes mariées constituant toujours un « problème », puisque certains les identifient par leur nom de mariée et d'autres par leur nom de naissance – pour supprimer les doublons d'une base de données à l'autre et mettre en place les relations (d'alliance, de filiation ou professionnelles) afin qu'elles soient intégrées dans la base *DEF19*.

Aujourd'hui, la base de données *DEF19*, développée sous Omeka S, se structure autour de deux entités : les individus et les établissements ou entreprises d'édition³⁸. Cela a conduit, lors de l'intégration des données, à quelques opérations de (re-)classification pour intégrer les données de la BNF, qui opte pour d'autres distinctions que celles choisies par *DEF19*. Typiquement, l'entreprise « Rey et Gravier », classée par la BNF comme une personne morale, représente pour *DEF19* : 1) une entreprise, 2) formée par deux individus. Compte tenu du fait que chaque entité présente un formulaire particulier, « Rey et Gravier » détient trois entrées dans *DEF19* : deux entrées individuelles, l'une à Pierre-Joseph Rey, l'autre à Jean-Simon Gravier, et une entrée relative à leur association. Le formulaire des entités « individus » est bien plus développé que celui relatif aux entreprises. De fait, le formulaire propre aux individus comprend neuf types d'informations, ou sous-formulaires, contre deux pour les établissements ou entreprises d'édition.

FIGURE 2. SCHÉMA DES FORMULAIRES INDIVIDUS DE LA BASE DE DONNÉES DEF19

Viera Rebolledo-Dhuin, avril 2019. DEF19, <http://def19.huma-num.fr>.

19

Lors du premier import massif de données, majoritairement issues de Libredit et de la BNF, et à côté de ces opérations de (re-)classification des entités dénommées, nous avons également procédé à une gestion des doublons créés par la fusion des diverses bases ressources. La suppression des doublons avant l'import massif a été réalisée par Viera Rebolledo-Dhuin de manière « manuelle », supposant la recherche de données biographiques (au sein de l'état civil) pour distinguer les homonymes, ainsi que par le maniement de formules et macros au sein d'un tableur basique. À l'avenir, et dans le cadre du deuxième import massif projeté, intégrant notamment Gutenberg, nous prévoyons d'utiliser, dans la mesure du possible, le logiciel OpenRefine³⁹, conseillé et utilisé par Maïwenn Bourdic.

FIGURE 4. ÉTAPES D'INTÉGRATION DES DONNÉES DANS LA BASE DEF19 (PREMIER IMPORT)

Viera Rebolledo-Dhuin

FIGURE 5. ÉTAPES D'INTÉGRATION DES DONNÉES DANS LA BASE DEF19 (DEUXIÈME IMPORT)

Viera Rebolledo-Dhuin

Plus précisément, il est arrivé que durant cette phase de dédoublement et d'identification, nous nous soyons affrontés à des conflits de contenus (dates d'état civil différentes selon les sources, par exemple). Malgré tout et à l'heure actuelle, le deuxième import est remis en question, compte tenu de notre objectif de publication papier du dictionnaire et du coût d'un nouvel import massif, que seule Limonade & Co peut actuellement gérer. De fait, l'entreprise nous a soutenus, dans la mesure du possible, en migrant notamment la base DEF19 vers une version stable d'Omeka S, en tenant compte de nos exigences. Cependant, les développeurs étant intervenus dans le cœur du programme, toute remontée vers des versions plus récentes d'Omeka S – intégrant les nouveaux plugins

(notamment d'import et d'export des données au format CSV, en vue de l'interopérabilité des données) – perturbe la visibilité de l'avancée du processus. Autrement dit, nous n'avons pas encore la main pour intégrer de nouvelles données, ni même pour modifier la structure de la base de données fondée principalement sur plusieurs ontologies généralistes et une ontologie plus spécifique :

- Bibliographic Ontology
- *DEF19* : concernant les brevets, contentieux, distinctions, etc.
- Dublin Core
- FOAF (*Friend of a Friend*)

Mise en place de collaborations scientifiques pour l'enrichissement des données à plus long terme. Le Web sémantique, une voie d'avenir ?

21 De fait, une deuxième étape d'intégration des données, sur laquelle Viera Rebolledo-Dhuin travaille actuellement, vise à prendre en compte les données de la base Gutenberg, celles issues du travail de Ian Maxted, celles provenant de l'Institut national de la propriété intellectuelle (INPI) et finalement celles de la base de l'Atelier data.bnf.fr. Plusieurs collaborations ont été mises en place au cours de ces deux dernières années afin d'enrichir, à plus long terme, la base *DEF19*. Aussi le Web sémantique favorise-t-il l'interopérabilité comme l'échange de données.

22 La base Gutenberg, répertoriant l'ensemble des brevetés français, nous a été livrée après l'intégration des données issues de la base Libredit. Gutenberg fonctionne par ailleurs quelque peu autrement que Libredit, notamment car les données biographiques et géographiques sont moins denses. Comme Libredit, elle présente cependant des liens avec les prédécesseurs et les successeurs des brevetés qu'il s'agit de préserver. Notre travail consiste donc à transformer le fichier texte en un fichier de données structurées tout en préservant les liens établis par les archivistes lors de la mise en place de cette base de données, ultérieurement transformée au format texte. Outre la restitution des liens relationnels, principalement entre prédécesseur et successeur, il s'agit également d'expurger la base Gutenberg de données déjà intégrées dans la base *DEF19* (comme celles issues de Libredit, qui comporte davantage de données qualitatives que Gutenberg, relativement aux brevetés parisiens). Par ailleurs, certains brevetés parisiens ont été – précédemment ou après leur expérience parisienne – brevetés en province, et font l'objet de fiches redondantes dans Gutenberg – qui n'a pas pleinement procédé au « nettoyage » des doublons. Le travail d'identification pour supprimer ces doublons tout en préservant les liens prédécesseur-successeur est donc d'autant plus long.

23 Suite à l'organisation de la première journée d'étude *DEF19*, en octobre 2015, Ian Maxted nous a par ailleurs soumis un fichier, issu de ses recherches personnelles, concernant 1 600 gens du livre normands ayant exercé entre 1480 et 1870. Les données structurées de cette base intégreront la base *DEF19* mais devront être passées au même crible que les données tirées de Gutenberg afin d'éviter les doublons tout en intégrant des données structurées sur le même format défini par l'équipe de recherche *DEF19*.

24 Au cours de ces deux dernières années, nous avons enfin pris des contacts, d'une part, avec l'INPI et, d'autre part, avec l'Atelier data.bnf.fr afin d'intégrer de nouvelles données structurées dans la base *DEF19*.

25 La rencontre avec les archivistes de l'INPI s'est opérée au lendemain de la journée d'étude consacrée aux dictionnaires des photographes, organisée par les AN en avril 2017 à l'occasion du mois de la photographie⁴⁰, et au sein de laquelle nous avons été invités à présenter notre projet et nos méthodes. De là est née une collaboration entre l'INPI et *DEF19* afin d'intégrer les éditeurs, détenteurs d'un brevet d'invention – conservés au sein de l'INPI – dans la base *DEF19*. L'intégration de ces données se fera dans un premier temps manuellement. Mais, il est à espérer que la mise en place de la nouvelle plateforme de l'INPI, en construction mais devant probablement se faire en Web sémantique, favorise les échanges de données RDF.

26 Viera Rebolledo-Dhuin a d'autre part, établi des contacts avec les membres de l'Atelier de data.bnf.fr dès 2015, l'idée étant de remédier à la transmission des notices nationales d'autorité de la BNF régulièrement mises à jour par le département de l'Inventaire rétrospectif. En effet, ce dernier nous communique à certaines échéances des lots de nouvelles notices concernant notre projet ainsi que des notices corrigées ou amplifiées. Or, la base *DEF19* ne cesse de s'enrichir parallèlement et comprend bien d'autres champs d'information que ceux compris dans les notices nationales d'autorité. Aussi, les imports massifs périodiques des notices corrigées par la BNF risquent d'écraser les données saisies directement sous Omeka S, sinon d'entraîner des doublons avec des notices créées précédemment au nouvel import. C'est pourquoi nous sommes en train de mettre en place une méthode pour automatiser la récupération des données de la BNF, *via* les *triplestores*⁴¹ de l'Atelier data.bnf.fr interrogeables sous SPARQL. Au-delà, et *DEF19* proposant des notices d'éditeurs non renseignées par l'inventaire rétrospectif de la BNF, il s'agit de collaborer activement à la rédaction des notices nationales d'autorité par la mise en place d'exports de champs validés par cette même BNF. La base de données *DEF19*, fondée sur la version sémantique d'Omeka, permet la réutilisation par l'Atelier data.bnf.fr et, par là même, le partage des données avec le département de l'Inventaire rétrospectif, chargé de la rédaction et de la validation des notices nationales d'autorité.

27 Autrement dit, un des enjeux de notre base s'avère de mettre en place un outil permettant d'échanger les données et de faire dialoguer des chercheurs et des institutions spécialisés dans l'histoire du livre.

Le Web sémantique au service d'une plateforme collaborative et ouverte pour l'analyse des données

28 Un des objectifs du programme *DEF19* consiste enfin, non pas seulement à recenser des données sur les éditeurs français du XIX^e siècle, mais également à construire une base collaborative, ouverte et permettant des analyses quantitatives variées.

Des collaborations scientifiques pour la visualisation et l'exploitation des données

29 Le développement de *DEF19* sous Omeka S vise à utiliser son potentiel en matière de visualisation et d'exploitation des données. De fait, Omeka est largement utilisé pour la mise en place d'expositions ou de bibliothèques virtuelles et propose de nombreux plug-ins de visualisation des données. Par ailleurs, le choix d'Omeka S favorise l'interopérabilité des données en vue de leur analyse quantitative. Des objectifs scientifiques et des choix techniques opérés, découlent donc diverses collaborations scientifiques et des pistes de valorisation de l'interface en cours de construction.

30 Omeka proposant un plugin le reliant à Zotero, nous avons d'abord mis en place une bibliographie collective sous ce logiciel, à laquelle nous avons ajouté, au format CSV, des références abrégées, sur le modèle des notices nationales d'autorité de la BNF, afin de les réimporter dans la base temporaire *DEF19* développée sous Omeka 2.4. Les abréviations des références bibliographiques doivent intégrer la base *DEF19* développée sous Omeka S, sous forme de liste déroulante, en bas de chacun des formulaires, pour faciliter la saisie des données tout en validant leur contenu.

31 En outre, afin de rendre la consultation de la plateforme *DEF19* plus agréable et de profiter des plugins existants sous Omeka, notre réflexion sur l'interface de consultation nous a conduits à établir un dialogue avec le groupe de recherche Geohistoricaldata (GHD⁴²). Celui-ci, dirigé par Maurizio Gribaudo (EHESS) et lié notamment à des chercheurs de l'Institut national de l'information géographique et forestière (IGN), propose une plateforme collaborative ouverte de géolocalisation historique à différentes échelles (France, Paris, Saint-Jean-de-Latran) du XVIII^e au XXI^e siècle. La collaboration ainsi mise en place entre *DEF19* et GHD vise à proposer aux utilisateurs de *DEF19*, au lieu d'un plugin de géolocalisation automatique des éditeurs français fondé sur l'ontologie de GeoNames – comportant plusieurs erreurs – et un fond de carte actuel de Google Maps sur lesquels s'appuie le plugin de géolocalisation d'Omeka, un plugin de géolocalisation historique construit de manière empirique, respectant ainsi les changements d'adresses (nombreux à Paris au XIX^e siècle), et fondé sur des fonds de cartes historiques (Cassini, Tardieu, Verniquet, Vasserot, Jacoubet, atlas municipal, etc.). L'interrogation de l'API (*Application Programming Interface* ou interface de programme applicative) proposée par GHD – ou du SIG (Système d'information géographique) historique que les chercheurs de GHD mettent à disposition –, comportant plusieurs couches spatiotemporelles, permettra de mieux saisir l'environnement géosocial de l'établissement des éditeurs français du XIX^e siècle. En sens inverse, *DEF19* participe à l'enrichissement de la base d'adresses, ce qui favorise le développement de GHD. À notre initiative, plusieurs réunions ont été organisées pour mettre en contact nos différents collaborateurs techniques et scientifiques (BU-UVSQ, Limonade & Co, département de l'Inventaire rétrospectif de la BNF, département Justice et Intérieur des AN, Atelier data.bnf.fr, INPI, GHD, etc.), au cours desquelles tous ont été convaincus

de la pertinence du partage des données relatives à la géolocalisation historique et à leur visualisation sur des fonds de cartes contemporains aux événements.

32 Enfin, dans le cadre de notre réflexion sur l'interface de consultation de la plateforme *DEF19*, et compte tenu des travaux de recherche en histoire quantitative des chercheurs participant à ce programme ANR, nous avons souhaité mettre en place d'autres plugins de visualisation des données et favoriser l'analyse quantitative de ces données par des chercheurs internes ou externes au programme. Cela suppose néanmoins de bénéficier de plugins d'export des données (structurées) proposés par Omeka et Omeka S dans des formats qui peuvent être supportés par d'autres logiciels d'analyse statistique ou d'analyse de relations sociales. Typiquement, Omeka propose un plugin d'export au format CSV (*comma-separated values*, format de texte tabulaire sous forme de valeurs séparées par une virgule), transférable dans d'autres logiciels d'analyse quantitative ou d'analyse de réseaux. Aussi, dans le cahier des charges livré à notre prestataire, nous avons demandé que les données soient exportables, avec ou sans filtre, sous différents formats, dont le format CSV, afin d'exploiter les données sous le logiciel multifonction R⁴³ ou des logiciels plus spécifiques, tel Puck⁴⁴, destiné à l'analyse des relations de parenté, mais, plus généralement, à celle des relations sociales. C'est dans ce cadre que nous menons une réflexion sur les plugins de visualisation des données relationnelles, en collaboration avec les membres du programme ANR Kinsources⁴⁵, pour enrichir la plateforme de consultation *DEF19*, en espérant que cela puisse aboutir à la mise en place de plugins de visualisation de type arbres généalogiques ou graphes de réseaux, pour rendre compte de l'environnement familial et social des éditeurs français du XIX^e siècle, et ce sur une page de notice individuelle d'éditeur.

33 Omeka S doit également proposer un plugin d'export des données RDF ou d'interrogation des triplestores de notre base de données à tout chercheur et toute chercheuse afin de leur permettre d'exploiter autrement les données collectées. Nous souhaitons en effet, depuis l'origine du projet, que notre plateforme soit pleinement ouverte : non seulement dans la récupération des données, mais aussi du point de vue de leur enrichissement par le biais d'une démarche collaborative.

DEF19, une plateforme collaborative en construction fondée sur le Web sémantique

34 L'adoption d'un logiciel fonctionnant avec les outils du Web sémantique offre de multiples avantages au programme ANR *DEF19*. Néanmoins, compte tenu du développement des logiciels fondés sur le Web sémantique, comme Omeka S, un certain nombre d'objectifs restent en construction. Trois chantiers nous préoccupent particulièrement : le premier concerne le lien entre *DEF19* et l'Atelier data.bnf.fr, le deuxième est lié à l'exploitation de l'API de GHD et le troisième, plus général, est lié à l'échange ou à la création collaborative de données.

35 Plusieurs fonctionnalités et plugins projetés n'ont pas pu être mis en place durant la première phase de développement de la plateforme *DEF19* par nos développeurs. De fait, à ce moment-là, Omeka S fonctionnait sous sa version alpha.

Bibliographie

- Arbour, Roméo. 2003. *Dictionnaire des femmes libraires en France (1470-1870)*. Genève : Droz.
- Barbier, Frédéric. 2002. *Lumières du Nord : imprimeurs, libraires et « gens du livre » dans le Nord au XVIII^e siècle (1701-1789)*. *Dictionnaire prosopographique*. Genève : Droz.
- Barbier, Frédéric, Sabine Juratic et Annick Mellerio. 2007. *Dictionnaire des imprimeurs, libraires et gens du livre à Paris, 1701-1789*, A-C. Genève : Droz.
- Bibliothèque nationale de France. 2004. *Répertoire d'imprimeurs/libraires (vers 1500 – vers 1810)*. Paris : Bibliothèque nationale de France.
- Devriès, Anik et François Lesure. 1979. *Dictionnaire des éditeurs de musique français. Des origines à environ 1820. Volume I*. Genève : Minkoff.
- Lefils-Boscq, Marie-Claire. 2013. « La Librairie parisienne sous surveillance (1814-1848) : imprimeurs en lettres et libraires sous les monarchies constitutionnelles ». Thèse de doctorat, université de Versailles–Saint-Quentin-en-Yvelines.
- Mellot, Jean-Dominique. 2006. « Qu'est-ce qu'un livre ? Qu'est-ce que l'histoire du livre ? Points de départ et perspectives ». *Histoire et civilisation du livre* 2 : 5-18.
- Mollier, Jean-Yves. 2008. « La police de la librairie, 1810-1881 ». Dans *Histoire de la librairie française*, édité par Patricia Sorel et Frédérique Leblanc, 16-26. Paris : Éditions du Cercle de la librairie.
- Préaud, Maxime, Marianne Grivel, Pierre Casselle et Corinne Le Bitouzé. 1987. *Dictionnaire des éditeurs d'estampes à Paris sous l'Ancien Régime*. Paris : Promodis et Éditions du Cercle de la librairie.
- Rebolledo-Dhuin, Viera. 2011. « La Librairie et le crédit : réseaux et métiers du livre à Paris, 1830-1870 ». Thèse de doctorat, université de Versailles–Saint-Quentin-en-Yvelines.
- Sorel, Patricia et Frédérique Leblanc, éd. 2008. *Histoire de la librairie française*. Paris : Éditions du Cercle de la librairie.
- Vaillant, Alain, éd. 1992. *Mesure(s) du livre. Colloque organisé par la Bibliothèque nationale et par la Société des études romantiques*. Paris : Bibliothèque nationale.
- Martin-de-Vesvrotte, Sylvie, Marie-Félicie Pérez et Henriette Pommier, éd. 2002. *Dictionnaire des graveurs-éditeurs et marchands d'estampes à Lyon aux XVII^e et XVIII^e siècles. Catalogue des pièces retrouvées*. Lyon : Presses universitaires de Lyon.
- Vouillot, Bernard. 1990. « La Révolution et l'Empire ; une nouvelle réglementation ». Dans *Histoire de l'édition française. Le livre triomphant, 1660-1830, tome II*, édité par Roger Chartier et Henri-Jean Martin, 694-708. Paris : Fayard et Éditions du Cercle de la librairie.

Notes

- 1 Cet article s'inspire de plusieurs présentations du projet DEF19 faites par Viera Rebolledo-Dhuin avec ou sans Jean-Charles Geslot ; voir notamment la conférence donnée le 14 mai 2018 aux *Lundis numériques de l'INHA* ([lien](#)) et celle du 17 janvier 2019 à l'Espace Mendès-France ([lien](#)).
- 2 Base de données (en construction) du [projet ANR Anticipation](#).
- 3 Bases de données du [Centre de recherche sur la littérature de voyage](#).
- 4 Plateforme Médias 19, base de données Mystères urbains.
- 5 Base de données du [projet ANR Littépub](#).
- 6 [Plateforme Médias 19](#), base de données Petite Presse.
- 7 École nationale des chartes, base de données [Esprit des livres](#).
- 8 Base de données du [projet ANR Mame & Fils](#).
- 9 Institut Coppet, [liste complète des titres publiés par Guillaumin \(1837-1910\)](#).
- 10 Groupe de recherches et d'études sur le livre au Québec, [base Catalogues d'éditeurs](#).
- 11 Dominique Varry, *Gens du livre à Lyon au XVIII^e siècle – notices biographiques*.

- 12 École nationale des chartes, *Dictionnaire des imprimeurs-lithographes du XIX^e siècle*.
- 13 Le Maitron. *Dictionnaire biographique. Mouvement ouvrier, mouvement social*, « Livre-papier-carton ».
- 14 Plateforme Médias 19, base de données Dictionnaire des journalistes.
- 15 Bibliothèque cantonale et universitaire de Lausanne, *Répertoire des imprimeurs et éditeurs suisses actifs avant 1800*.
- 16 Groupe de recherches et d'études sur le livre au Québec, *base de données Gens du livre au Québec*.
- 17 *Biographical and Bibliographical Information on the Book Trades*.
- 18 Biblioteca virtual Miguel de Cervantes, *Editores y editoriales iberoamericanos (siglos XIX-XXI)*.
- 19 Le Web sémantique, ou Web de données, constitue une extension du Web, standardisée par le W3C (*World Wide Web Consortium*) – principal organisme de normalisation du Web –, facilitant l'exposition, l'exploitation et l'échange de données structurées au format RDF (*Resource Description Framework*).
- 20 Sont cependant exclus les éditeurs qui n'ont publié que des partitions de musique, des atlas ou des lithographies (il s'agit, en effet, le plus souvent d'un travail d'impression spécifique plus que d'édition, et certaines de ces catégories ont déjà fait l'objet d'entreprises de recensements), ainsi que les éditions à compte d'auteur (qui ne concernent donc pas des professionnels de l'édition).
- 21 Le décret du 5 février 1810, confirmé par la loi du 21 octobre 1814, impose la détention d'un brevet aux imprimeurs typographes et aux libraires. Cette obligation est étendue aux imprimeurs lithographes par l'ordonnance du 8 octobre 1817 puis, par le décret du 22 mars 1852, aux imprimeurs en taille-douce.
- 22 Si l'« éditeur » n'apparaît que « sous les plis de la loi » en 1810 (Mollier 2008), il est expressément nommé dans le décret de 1852. Ainsi, de célèbres libraires-éditeurs – déjà bien établis, mais non brevetés –, comme Michel Lévy ou Gervais Charpentier, ne font la demande de brevet qu'à cette époque.
- 23 Ont participé, depuis 2014, et à des titres divers, à la partie scientifique de DEF19 (définition du projet, conception de la base, collecte des données) : au CHCSC, Marie-Claire Lefils-Boscq, Nicole Dufournaud (IGE), Jean-Charles Geslot, Nicolas Mellot (stagiaire), Jean-Yves Mollier, Patricia Sorel (également pour le pôle Métiers du livre), Viera Rebolledo-Dhuin (post-doctorante) ; à la BU de Saint-Quentin, Renaud Delemontez, William Demet, Amin Galib Ali, Claire Lebreton, Aleth Tisseau Des Escotais ; à la BNF, Nathalie Aguirre, Rémi Baratault (stagiaire), Jean-Dominique Mellot ; aux Archives nationales, Maïwenn Bourdic, Nicole Brondel, Violaine Challéat-Fonck, Denise Ogilvie, Marion Veyssière. Le projet a également bénéficié, au sein du CHCSC, de l'accompagnement administratif, financier, documentaire et scientifique de Céline Clouet, Aurélia Desplains, Mélodie Hack-Passemar, Esther Van Rossen, Hélène Veillard. Il a enfin profité du concours technique de Pascal Cristofoli et Gérard Martinelli.
- 24 Un compte rendu de cette journée d'étude a été publié sur le *carnet de recherche DEF19* et renvoie aux diverses captations vidéo réalisées à l'occasion.
- 25 <http://www.qwamci.com/qes-infomedia-suite/>.
- 26 <https://www.huma-num.fr/>.
- 27 <https://www.accesstomemory.org/fr/>.
- 28 <https://www.ica.org/fr>.
- 29 <http://archives.mae.u-paris10.fr/>.
- 30 <http://omeka.org/classic/>.
- 31 <https://rrchnm.org/>.
- 32 Un même individu peut être breveté imprimeur et/ou libraire, dans un même lieu ou non, de manière successive ou simultanée, à une seule ou à plusieurs reprises (selon l'évolution de la législation, notamment). Il est donc nécessaire de créer un formulaire spécifique à chaque brevet en établissant un bloc de champs répétables, soit un ensemble de champs qui ne se superposent pas l'un l'autre, mais qui rendent compte de la cohérence de chaque brevet (lieu, type de brevet, date de début et de fin de brevet, prédécesseur ou successeur, etc.).
- 33 <http://www.limonadeandco.fr/>.

34 <http://www.limonadeandco.fr/interoperabilite-entre-omeka-et-nakala-nakalona/>.

35 <https://omeka.org/s/>.

36 Viera Rebolledo-Dhuin tient, à ce propos, à remercier chaleureusement M^{me} Choublier-Grimbert, qui a assuré la vacance de poste de M. Patrick Laharie, pour cette transmission qui lui a été fort utile.

37 Plusieurs outils ont été mis en ligne, en septembre 2018, par Maïwenn Bourdic sur le site des Archives nationales : [Brevets des imprimeurs, libraires et lithographes \(Paris et Seine\) : dossiers nominatifs \(1810-1881\)](#), F/18/1726-F/18/1837 ; [Brevets des imprimeurs, libraires et lithographes des départements \(hors Paris et Seine\) : dossiers nominatifs \(1810-1881\)](#) ; Liste générale des brevetés de l’Imprimerie et de la Librairie (Premier Empire, Restauration) F/18(I)/14 à F/18(I)/16, F/18(I)/18, F/18(I)/22 et F/18(I)/24, qui concerne la France mais également les départements du Premier Empire (aujourd’hui en Allemagne, Belgique, Italie, Pays-Bas, Suisse).

38 Dans une précédente étape de développement, il existait également une entité « organisations », qui visait à faciliter la saisie des éléments d’inscription sociale des individus et des entreprises d’édition. Compte tenu de son incohérence scientifique pour notre base de données, cette entité a disparu et la participation à des associations ou autres organisations collectives figure dans les fiches individus ou entreprises sans que le nom de l’organisation constitue une entité à part entière.

39 Le logiciel OpenRefine constitue la nouvelle version de Google Refine et permet de nettoyer et mettre en forme des données de manière massive.

40 Cette journée d’étude, annoncée sur le site des Archives nationales, a fait l’objet de captations vidéo. La communication consacrée à DEF19 est visible au lien suivant : <https://www.dailymotion.com/video/x5mbl42>.

41 Le *triplestore*, ou *RDF store* – le RDF étant le langage de base du Web sémantique –, est une base de données pouvant stocker des milliards de données, d’un seul type – le triplet RDF, structuré sous la forme « sujet, prédicat, objet » –, et en récupérer autant à l’aide du langage de requête SPARQL.

42 Voir le site [geohistoricaldata](#), et la [présentation](#) faite le 24 mars 2017.

43 Permettant à la fois l’analyse statistique et de réseaux, voir au lien suivant : <https://www.r-project.org/>.

44 Présentation sur le site [Kinship et Computing](#).

45 <https://www.kinsources.net/>.

Auteurs

Jean-Charles Geslot

EA 2448 CHCSC, université de Versailles–Saint-Quentin-en-Yvelines, Guyancourt, France
Jean-Charles Geslot est historien, maître de conférences en histoire contemporaine, spécialiste de l’histoire du livre et de l’édition au XIX^e siècle, et notamment de l’édition historique. Membre du Centre d’histoire culturelle des sociétés contemporaines (EA 2448), il coordonne le projet DEF19 (*Dictionnaire des éditeurs français du XIX^e siècle*).
jean-charles.geslot@uvsq.fr

Viera Rebolledo-Dhuin

EA 2448 CHCSC, université de Versailles–Saint-Quentin-en-Yvelines, Guyancourt, France
Viera Rebolledo-Dhuin, docteure en histoire contemporaine, est spécialiste de l’histoire du livre et de l’édition au XIX^e siècle, de l’histoire relationnelle du crédit au sein des métiers du livre et des méthodes quantitatives pour historiens. Elle est membre du Centre d’histoire culturelle des sociétés contemporaines (EA 2448).
viera.rebolledo-dhuin@uvsq.fr

Droits d'auteur

Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](#).