



**HAL**  
open science

## De l'index papier à la base de données et aux cartes : les registres médiévaux du Trésor des chartes

Virgile Reignier, Dominique Stutzmann

### ► To cite this version:

Virgile Reignier, Dominique Stutzmann. De l'index papier à la base de données et aux cartes : les registres médiévaux du Trésor des chartes. Les Amis de l'IRHT. Bulletin de l'association, 2022, pp.2-4. hal-04195105

**HAL Id: hal-04195105**

**<https://hal.science/hal-04195105>**

Submitted on 4 Sep 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

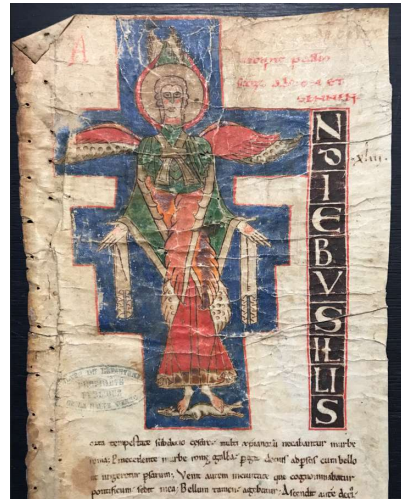
en effet, sujet à un renouvellement périodique, qui aboutissait à la copie de nouveaux volumes enregistrant les mises à jour du culte ; les vieux livres, devenus inutiles à la pratique liturgique, étaient jetés ou plutôt réutilisés de la manière la plus disparate, souvent en tant que support pour relier d'autres volumes (couvrures, gardes, contregardes, etc.). Au fil des siècles, les liturgies locales ont été effacées pour laisser place à un rituel plus uniforme et moins caractérisé. La liturgie de Rome s'impose, surtout à partir du XIII<sup>e</sup> siècle, grâce à l'adoption du Missel et du Bréviaire romains par les Franciscains. C'est donc dans les fragments qu'on retrouve la seule trace d'une liturgie locale ou d'un culte particulier, et c'est seulement à travers le repérage et l'étude des fragments qu'on pourra reconstituer le patrimoine liturgique et musical du Moyen Âge latin, du moins en partie.

Le projet, qui a démarré en 2021, se propose d'enquêter sur les principaux dépôts parisiens conservant des fragments de manuscrits médiévaux : la Bibliothèque nationale de France, la Bibliothèque Mazarine, la Bibliothèque Sainte-Geneviève, la Bibliothèque de la Ville de Paris, la Bibliothèque de la Sorbonne et les Archives nationales. Le projet s'inscrit également dans la continuité d'un autre projet, « Tracing the Past », réalisé dans le cadre de *Fragmentarium* de l'université de Fribourg (2019-2020), qui avait étudié une bonne partie des fragments détachés de la BnF.

L'année 2021 a été consacrée à la consultation des fragments conservés aux Archives nationales à Pierrefitte-sur-Seine : il s'agit de 25 boîtes provenant des défects de reliures que les Archives départementales de France avaient envoyés dans les années 1920 aux Archives nationales, pour un total d'environ 2000 pièces. Le fonds (AB XIX 1722-1746) est organisé par département ; à l'intérieur de chaque boîte, en revanche, les fragments ne sont pas numérotés, ce qui a rendu le référencement difficile. L'atelier de restauration des AN a entrepris une action de dépoussiérage et inventarisé de chaque pièce, qui, pour l'instant, ne s'est concrétisée que pour les deux premières boîtes.<sup>1</sup>

Le contenu de chaque boîte est assez hétérogène : aux fragments liturgiques se mêlent à des fragments juridiques, patristiques, hagiographiques et littéraires. Certains ont été analysés dans des études ponctuelles en raison de leur intérêt historique, mais un travail systématique d'identification reste à faire. Les fragments liturgiques portant des notations musicales, qui feront l'objet d'une prochaine publication, ont été en partie étudiés par Solange Corbin (*La notation musicale neumatique. Les quatre provinces Lyonnaises : Lyon, Rouen, Tours et Sens*, thèse, EPHE, Paris 1957) ; elle les avait ensuite signalés à Michel Huglo. C'est par une mention de ce dernier (« Le domaine de la notation bretonne », *Acta musicologica*, 35, 1963, p. 54-84) qu'on a pu retracer les membra disiecta d'un antiphonaire-graduel de la fin du XII<sup>e</sup> siècle (ou début du XIII<sup>e</sup>) conservés dans quatre dépôts différents, à savoir les AN (boîte AB XIX 1742), les Archives départementales de Maine-et-Loire, les Archives départementales d'Indre-et-Loire et l'abbaye Saint-Pierre de Solesmes : la vérification des originaux a confirmé la dispersion de ce manuscrit et a ainsi fourni de nouvelles pistes de recherche pour de possibles provenances.

Ces derniers fragments sont, à notre connaissance, les seuls à transmettre l'Office de saint Hermeland d'Indre et leur contenu confirme leur importance. En effet, il n'est pas rare de trouver des pièces inconnues ou très peu attestées par ailleurs, surtout dans les offices du sanctoral. Dans un groupe de fragments provenant du département de la Vienne on découvre des répons inédits pour la *reversio sanctae Radegundis* ; un bifeuillet d'un bréviaire monastique provenant de la Côte-d'Or nous dévoile un office incomplet pour la translation de saint Nicolas dont aucun autre témoin n'a été repéré jusqu'à présent ; un fragment du département de la Haute-Vienne nous offre le début de la passion des saints Abdon et Sennen, accompagné d'une superbe représentation d'un séraphin piétinant une bête.



Paris, AN, AB XIX 1745, début de la passion des saints Abdon et Sennen.

La plupart des textes liturgiques sont malheureusement mutilés, comme souvent, et la reconstitution de la liturgie, telle qu'elle était célébrée en différents endroits et dans son développement chronologique, semble une tâche utopique. Mais, comme un restaurateur d'art peut arriver à donner une idée plutôt précise d'une fresque à partir des images et des figures fragmentaires qu'il découvre sur un mur, ces fragments de livres liturgiques, une fois récupérés et contextualisés, constitueront les pièces d'un puzzle qui, à l'aide d'un effort de grande envergure, pourra redonner vie à la liturgie médiévale.

## De l'index papier à la base de données et aux cartes : les registres médiévaux du Trésor des chartes

Virgile REIGNIER  
*Doctorant à l'École nationale des chartes*  
 Dominique STUTZMANN  
*Section de paléographie latine*

Le projet de recherche européen Himanis, dirigé par l'IRHT et ayant mené à la création d'un moteur de recherche en plein texte sur les 80 000 pages des registres du Trésor

<sup>1</sup> La plupart de ces fragments ont été numérisés (2 500 prises de vues) et mis en ligne sur la BVMM par Gilles Kagan en 2021-2022.

des chartes (Paris, Archives nationales, JJ 35-JJ 211) est achevé depuis 2017. Les travaux sur ce corpus se poursuivent néanmoins à l'IRHT dans d'autres cadres tels que Biblissima+ et le projet européen HOME (History of medieval Europe).

Les travaux en cours concernent l'amélioration des technologies pour identifier les noms de personnes et de lieux dans le texte des actes de la chancellerie royale et pour proposer des identifications précises et désambiguïsées à partir d'une transcription automatique par ordinateur. Virgile Reignier, étudiant de l'École nationale des chartes, a exploité l'index de l'inventaire des registres JJ 37-50 couvrant les années 1303 à 1314<sup>1</sup>. Il l'a transformé en un référentiel structuré et géolocalisé pour entraîner l'intelligence artificielle qui procédera à l'identification des noms de lieux.

### Expliciter les structurations implicites

L'index d'origine se présente sous la forme d'une table pour les sujets et d'une autre pour les personnes et les lieux. Numérisé en même temps que l'inventaire puis traité par OCR et structuré dans un document XML conforme aux spécifications de la Text Encoding Initiative, il forme un référentiel contenant 4229 sujets, 7795 noms de lieux et 6534 noms de personnes.

La grande qualité de l'index repose sur une structuration très fine avec des renvois multiples où la structuration est fournie par des mises en pages, signes de ponctuation et abréviations, portant tous une part d'ambiguïté. La structure est classique... et complexe :

1. *Forme simple*. Entrée d'index suivie des numéros d'entrée d'inventaire correspondants.
2. *Forme subordonnée*. Par exemple : « Val-des-Écoliers (Ordre du) : [...], — chapitre général ».
3. *Forme rejetée*. Entrée disposant de renvois vers une ou plusieurs autres entrées. Par exemple : « Abandons, v. donations, renonciations, remises ». Chacun des termes « donations », « renonciations » et « remises » constitue un renvoi vers ces mots disposant de leur entrée propre.
4. *Forme subordonnée à une forme rejetée*. Par exemple, sous la même entrée principale : « Abandons, v. donations, renonciations, remises, [...] — spontanés au roi », le tiret cadratin introduit l'entrée « [Abandons] spontanés au roi », qui est une entrée subordonnée à l'entrée « Abandons ».
5. *Forme subordonnée rejetée*. Entrée sous forme subordonnée disposant de renvois vers une ou plusieurs autres entrées associées. Par exemple : « Baux, [...], — à rente, v. arrentements ». Le renvoi vers l'entrée « arrentements » ne concerne que la sous-entrée « [Baux] à rente ».
6. *Double subordination*. Entrée subordonnée à une entrée elle-même sous forme subordonnée. Par exemple : « Vidimus : [...], — et confirmations de Philippe le Bel [...], — de ses prédécesseurs » désigne les vidimus et confirmations donnés par Philippe le Bel pour des actes donnés par les rois qui l'ont précédé et non pas des copies insérées de vidimus et confirmations antérieures. Le phénomène concerne aussi les noms de lieux et, par exemple, l'entrée « Paris » présente des subordinations multiples, dont un niveau implicite par paragraphe

(Paris > [églises] > église cathédrale de Notre-Dame > chanoines).

7. *Forme rejetée subordonnée à une forme rejetée*. Entrée disposant de renvois vers une ou plusieurs entrées associées et subordonnées à une entrée elle-même sous forme rejetée.

Donations, dons : v. aussi legs; --- à charge d'hommage, v. fiefs-rentes, inféodations; — à titre précaire, 268, 576, 1955; — au roi, 2221; v. aussi legs au roi.

Face à ces enchevêtrements multiples, de nombreuses entrées ont dû être corrigées manuellement.

### Les noms de lieux

Une seconde phase du travail a porté spécifiquement sur les noms de lieux pour les associer à des coordonnées géographiques. Les référentiels utilisés pour l'alignement automatique sont GeoNames ([www.geonames.org/](http://www.geonames.org/)) et le *Dictionnaire topographique de la France* ou « DicoTopo » ([dicotopo.cths.fr/](http://dicotopo.cths.fr/)). L'alignement automatique a été complété par un traitement manuel en raison des fautes d'orthographe, erreurs d'OCR ou de changement de nom des communes.

Les entrées sont présentées selon des normes classiques : toponyme puis, entre crochets, département, canton (mais non arrondissement), commune. Ces normes posent néanmoins des problèmes aux traitements automatiques : informations implicites, inversion des particules, compléments de noms composés (par ex. « Abergement (L') [Saône-et-Loire, c<sup>on</sup> Buxy, c<sup>nc</sup> Messey-sur-Grosne] » ou « Chilly [-Mazarin, Seine-et-Oise, c<sup>on</sup> Longjumeau] »). Pour les entités incluses dans une commune, on a souvent recouru aux coordonnées de la commune.

Certains toponymes n'ont pas pu être associés à une entité externe unique et géolocalisée. Auquel cas, le toponyme simple est enregistré, et lié à chacune des propositions pouvant lui correspondre. Les zones situées à cheval sur plusieurs entités géolocalisables ont été alignées avec les coordonnées de chacune de ces entités et l'on a proposé des coordonnées moyennes. Outre les noms non localisables, on a souvent dû corriger :

1. *Nom donné dans l'index, identifications multiples automatiques, dont certaines fautives*. Par exemple : Beaumont [Loiret, c<sup>on</sup> Beaugency, c<sup>nc</sup> Cravant], associé automatiquement à deux « Beaumont » dans le canton de Beaugency (code Insee 45116, GeoNames 12307408 / code Insee 45270, GeoNames 12307422). Le second est fautif.
2. *Nom donné dans l'index, plusieurs solutions possibles, mention d'incertitude dans l'index*. Par exemple : « Abbeville [Seine-Maritime, région de Bacqueville-en-Caux ou d'Offranville ?] », associé automatiquement à quatre « Abbeville » en Seine-Maritime (code Insee 76050, DicoTopo P45419125 ; code Insee 76659, DicoTopo P65099007 ; code Insee 76255, DicoTopo P65312883 ; code Insee 76051, DicoTopo P86515984). Le quatrième correspond strictement à l'entrée d'inventaire, puisqu'il s'agit précisément d'un toponyme

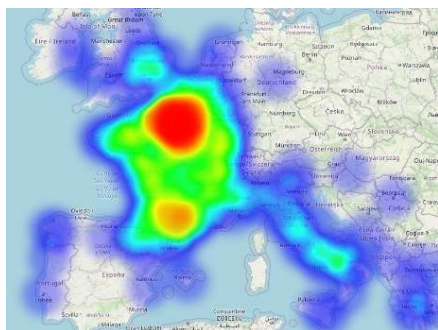
<sup>1</sup> Jean GLÉNISSON et Jean GUÉROUT, *Registres du trésor des chartes: inventaire analytique*, 1: Règne de Philippe le Bel, sous la dir. de Robert FAWTIER, Paris, 1958.

d'identification incertaine correspondant à la formulation de l'index.

3. *Nom donné dans l'index, plusieurs solutions possibles sauf argument e silentio.* Par exemple : Avesnes (Nord), associé à Avesnes-sur-Helpe (code Insee 59036), Avesnes-les-Aubert (code Insee 59037), Avesnes-le-Sec (code Insee 59038). Seule l'absence d'indication de canton permet d'identifier Avesnes-sur-Helpe, chef-lieu de canton, dans l'indexation de la locution « Hugues de Châtillon, comte de Blois et sire d'Avesnes ». Or l'argument par absence est toujours difficile à manier dans les traitements automatiques, pis encore après le redécoupage cantonal de 2014.

### Cartographie des résultats

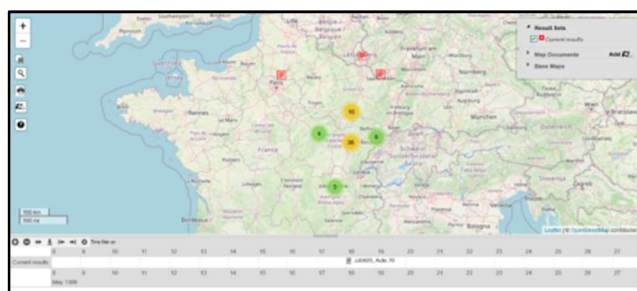
Notre base de données comprend désormais les lieux, les liens entre eux, ainsi que les renvois aux actes concernés. Elle permet de créer une carte dynamique pour visualiser les lieux présents dans les registres JJ 37-JJ 50 et pondérés selon le nombre d'actes les mentionnant (<https://virgile-reignier.github.io/Carte-JJ37-50/>) :



Carte générale

Cette carte propose une vision de l'implantation géographique des intérêts et de l'action du roi entre 1303 et 1314. Elle s'inscrit dans la lignée de travaux récents portant sur le déploiement de l'autorité du roi à la fin du Moyen Âge (par ex. L. Dauphant) et les complète pour le début du XIV<sup>e</sup> siècle. Ce travail pourra être enrichi par la suite par l'incorporation des données contenues dans les autres instruments de recherche du Trésor des chartes et par l'amélioration de la caractérisation des lieux mentionnés.

Si les spécialistes de l'espace politique et les historiens locaux exploitent déjà ces informations, une telle visualisation fournit un point d'entrée nouveau pour les historiens moins spécialisés. De plus, l'intégration des toponymes accompagnée d'un rattachement aux actes à travers le numéro d'inventaire, permet aussi de voir les lieux concernés acte par acte.



Carte des toponymes mentionnés dans l'acte n° 76 du registre Paris, Arch. Nat., JJ42B.

Une fois le modèle complété, l'objectif final du projet est d'utiliser les données de l'index comme matériel

d'entraînement pour le déploiement de modèles de liage d'entités. Des techniques de « Reconnaissance des entités nommées » sont aujourd'hui déjà capables de repérer les noms de lieux, de personnes et de sujets dans les archives anciennes. Une fois ces éléments repérés, le liage avec les entités de la base de données permettra d'approfondir cette technique, puis de l'appliquer aux registres du Trésor des chartes qui ne disposent pas d'index.

Ce travail fait l'objet d'un mémoire de master :

Virgile REIGNIER, *Vers l'indexation automatique du Trésor des chartes. Constitution, alignement et utilisation d'un référentiel d'entités nommées au sein du projet Himanis*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Dominique Stutzmann et Thibault Clérice, École nationale des chartes, 2022.

## Les chemins de l'écriture savante au Sahara et au Sahel (XV<sup>e</sup>-XIX<sup>e</sup> siècle)

Ismail WARSCHIED

Section arabe

De nos jours, Tombouctou apparaît comme le symbole par excellence d'une culture de l'écrit propre aux peuples africains. Ce qui est moins connu, c'est que l'essor de la ville comme foyer d'érudition musulmane participe d'un processus plus vaste qui concerne l'ensemble de la zone s'étendant du fleuve Niger aux bords de l'Atlantique. Entre le XV<sup>e</sup> et le XIX<sup>e</sup> siècle, les traditions savantes de l'Islam s'enracinent dans ce *Far West* de l'Islam où mondes araboberbères et subsahariens se rencontrent. Le processus marque une nouvelle phase dans l'islamisation qui, jusque-là, est resté relativement restreinte. De façon encore plus importante, il articule la diffusion à large échelle de pratiques culturelles, intellectuelles et normatives au cœur desquelles se retrouve l'usage de l'écrit. Sahéliens et Sahariens s'attachent à composer des textes s'inscrivant dans le *curriculum* des sciences islamiques postérieur à 1300 : des commentaires, des chroniques, des récits de voyages, des traités de droit et de théologie, des poèmes et des prières. Mais ils se mettent aussi à échanger des lettres, à rédiger des actes notariés, à griffonner des formules pour conjurer l'invisible. Évoquer les manuscrits sahélo-sahariens signifie recourir à une impressionnante production textuelle tant savante que pragmatique, dont nous commençons seulement à entamer l'exploration. Les vestiges se trouvent dans des collections privées et publiques des pays de la région : la Mauritanie, le Mali, le Niger, le Nigéria, mais aussi l'Algérie et le Maroc, auxquelles s'ajoutent des fonds « exilés », à commencer par la bibliothèque du cheikh soufi et leader politique al-Hājj 'Umar Tāl (m. 1864), conservée à la BnF.

Le programme de recherche que je réalise depuis 2015 à la section arabe de l'IRHT interroge l'héritage textuel des peuples sahélo-sahariens dans une double perspective. Il s'agit de comprendre, tout d'abord, comment la pratique des savoirs islamiques s'articule socialement et culturellement. Comment contribue-t-elle à façonner les échanges sociaux, les relations de pouvoir ainsi que les discours qui les soutiennent, avant l'avènement de la colonisation ? Dans quelle mesure la diffusion de modèles culturels et normatifs provenant de contextes historiquement et géographiquement