



**HAL**  
open science

## Des chartae au Corpus: La plateforme des CBMA - Chartae/Corpus Burgundiae Medii Aevi

Eliana Magnani

► **To cite this version:**

Eliana Magnani. Des chartae au Corpus: La plateforme des CBMA - Chartae/Corpus Burgundiae Medii Aevi. Christelle Balouzat-Loubet. Digitizing Medieval Sources. Challenges and Methodologies /L'édition en ligne de documents d'archives médiévaux. Enjeux, méthodologie et défis, 27, Brepols, pp.57-67, 2020, Atelier de Recherches sur les Textes Médiévaux, 978-2-503-58413-3. 10.1484/M.ARTEM-EB.5.117328 . hal-02510502

**HAL Id: hal-02510502**

**<https://hal.science/hal-02510502v1>**

Submitted on 17 Mar 2020

**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.

Des *chartae* au Corpus. La plateforme des CBMA - *Chartae/Corpus Burgundiae Medii Aevi*

Eliana MAGNANI

LAMOP (UMR 8589 CNRS)

Paru dans *Digitizing Medieval Sources. Challenges and Methodologies /L'édition en ligne de documents d'archives médiévaux. Enjeux, méthodologie et défis*, éd. Christelle BALOUZAT-LOUBET, Turnhout, 2019 (Atelier de Recherches sur les Textes Médiévaux, 27), p. 57-67 - <https://doi.org/10.1484/M.ARTEM-EB.5.117328> - ISBN 978-2-503-58413-3 / eISBN 978-2-503-58414-0

Les CBMA sont une plateforme documentaire électronique fondée sur la numérisation de documents diplomatiques édités et de manuscrits (cartulaires) relatifs à la Bourgogne médiévale. Le corpus évolue actuellement vers l'incorporation d'autres types documentaires. À la fin de l'année 2004, lorsque le programme démarre, la formule aujourd'hui consacrée des *Digital Humanities* venait juste d'éclorre<sup>1</sup>. Le projet s'inscrivait toutefois dans le sillage d'une série d'initiatives qui, depuis la fin des années 1960, visaient à intégrer l'informatique dans l'atelier du médiéviste<sup>2</sup>, et que les évolutions technologiques ainsi que le développement d'Internet au tournant du XXI<sup>e</sup> siècle ont définitivement institué dans l'habitus de l'historien<sup>3</sup>.

---

<sup>1</sup> Le terme *Digital Humanities* serait apparu en 2004 à la place de *Humanities Computing* dans S. SCHREIBMAN, R. SIEMENS, J. UNSWORTH (éd.), *A Companion to Digital Humanities*, Oxford, 2004, URL : <<http://www.digitalhumanities.org/companion/>> (consulté le 30 mai 2017). Sur la constitution de ce champ disciplinaire on peut se rapporter à M. K. GOLD, « The Digital Humanities Moment », *Debates in the Digital Humanities*, éd. M. K. GOLD, Minneapolis-London, 2012, p. IX-XVI et L. F. KLEIN, M. K. GOLD, « Digital Humanities: The Expanded Field », *Debates in the Digital Humanities*, éd. M. K. GOLD, L. F. KLEIN, Minneapolis-London, 2016, les deux volumes sont disponibles en ligne dans <<http://dhdebates.gc.cuny.edu/debates>> (consulté le 30 mai 2017).

<sup>2</sup> En France, dans le domaine de l'histoire médiévale et de la diplomatique en particulier, il faut citer des projets notables comme la création en 1966 à l'université de Nancy II du « Centre de recherches et d'applications linguistiques (C.R.A.L.) » qui aboutira en 2010 à la publication en ligne des « Chartes originales antérieures à 1121 conservées en France », ou la parution de la revue « Le médiéviste et l'ordinateur » entre 1979 et 2006. Cf. <<http://www.cn-telma.fr/originaux>> et <<http://www.menestrel.fr/spip.php?rubrique667>> (consultés le 30 mai 2017). Pour une synthèse raisonnée de ce moment historiographique, voir N. PERREAUX, *L'écriture du monde. Dynamique, perception, catégorisation du mundus au Moyen Âge (VII<sup>ème</sup>-XIII<sup>ème</sup> siècles)*. Recherches à partir de bases de données numérisées, Thèse Université de Bourgogne, Dijon, 2014, p. 100-113.

<sup>3</sup> N. DELALANDE, J. VINCENT (éd.), *Le métier d'historien à l'ère numérique : nouveaux outils, nouvelle épistémologie ?*, *Revue d'histoire moderne et contemporaine – supplément*, n°58-4bis (2011/5) ; F. CLAVERT, J. DANIEL, H. FLECKINGER, M. GRANDJEAN, F. IDMHAND, « Histoire et humanités numériques : nouveaux terrains de dialogue entre les archives et la recherche », *La Gazette des Archives*, Association des archivistes

En revenant sur l'expérience des CBMA et sur ses accomplissements, sans taire ses difficultés, cet article entend contribuer à la réflexion sur le processus de la publication en ligne et de la multiplication de la numérisation des documents médiévaux ainsi que sur leurs enjeux pour la recherche. Il s'agira dans un premier temps de rappeler l'histoire et le contexte institutionnel du projet, ensuite de détailler ses choix techniques et épistémologiques, de même que ses dernières évolutions.

## Un peu d'histoire institutionnelle

Le programme des CBMA est le fruit d'une double conjoncture favorable : la mise en place de projets numériques nationaux successifs et de compétences professionnelles décisives. L'idée de numériser des documents diplomatiques bourguignons m'a été proposée par Benoît-Michel Tock en 2004, en m'invitant à participer au projet intitulé « Chartes et pouvoir au Moyen Âge » (2004-2007) qu'il mettait en place dans le cadre de l'ACO (Appel à Contrat d'Objectifs) « Réseau des Maisons des Sciences de l'Homme - Terrains, technique, théories ». Réunissant différents partenaires nationaux, ce projet entendait numériser des actes édités antérieurs à 1300, et il a été poursuivi ensuite au cours de deux projets thématiques « Corpus » de l'ANR (Agence Nationale de la Recherche), ESPACHAR (« Les espaces de la charte » - 2007-2011) et CHARCIS (« Chartes cisterciennes du XII<sup>e</sup> siècle » - 2013-2016), qui se tournait vers la transcription d'actes inédits<sup>4</sup>. Le début de cette entreprise collective, qui a abouti à la réalisation de la base de données *Chartae Galliae*<sup>5</sup>, coïncidait avec l'arrivée de Marie-José Gasse-Grandjean dans l'équipe de l'UMR Arhehis à Dijon, où j'étais moi-même affectée depuis cinq ans en tant que chercheuse au CNRS, et au sein de laquelle le programme s'est développé jusqu'en 2014<sup>6</sup>. Ingénieure au CNRS, ayant travaillé auparavant à Nancy au sein de l'Artem (Atelier de recherche sur les textes médiévaux) dans l'important projet des « Chartes originales antérieures à 1121 conservées en France », M.-J. Gasse-Grandjean, forte de cette expérience et consacrant une grande partie de son temps de travail aux CBMA, a pu donner l'élan nécessaire à ce programme depuis ses débuts et jusqu'en 2014<sup>7</sup>. Au long de cette période le projet a rassemblé un réseau de collaborateurs extérieurs, a accueilli des post-doctorants<sup>8</sup> et a employé temporairement entre 2008 et 2010 plusieurs étudiants<sup>9</sup>, dont deux,

---

français, 2017, *Meta/morphoses. Les archives, bouillons de culture numérique*, 245 (1), p. 121-134 <halshs-01521814>.

<sup>4</sup> B.-M. TOCK, « Chartes et pouvoir au Moyen Âge », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 9, 2005, p. 173-177, disponible en ligne sur <<http://cem.revues.org/747>> ; DOI : 10.4000/cem.747; Id., « Le projet CHARCIS – Chartes cisterciennes du XII<sup>e</sup> siècle », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 17.2 (2013), disponible en ligne sur <<http://cem.revues.org/13189>> ; DOI : 10.4000/cem.13189 (consultés le 30 mai 2017).

<sup>5</sup> URL : <<http://www.cn-telma.fr/chartae-galliae>> (consulté le 30 mai 2017).

<sup>6</sup> UMR 5594, puis 6298, Arhehis (Archéologie, terre, histoire, sociétés - CNRS - Université de Bourgogne - Ministère de la Culture). Cf. E. MAGNANI et M.-J. GASSE-GRANDJEAN, « CBMA - *Chartae Burgundiae Medii Aevi*. Les débuts du projet », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 9 (2005), p. 178-179, disponible sur <<http://cem.revues.org/751>> ; DOI : 10.4000/cem.751 (consulté le 30 mai 2017).

<sup>7</sup> Voir les différents articles, études et comptes rendus des journées d'études rassemblés dans la « Collection CBMA », disponible en ligne sur <<http://cem.revues.org/12172>> (consulté le 30 mai 2017).

<sup>8</sup> Isabelle Rosé a bénéficié d'un post-doctorat CNRS (2006-2007), tandis que Marlène Helias-Baron a travaillé comme ingénieure de recherche dans le cadre de l'ANR CHARCIS (2013-2014).

<sup>9</sup> Lucie Hoornaert, Benjamin Lefèvre, Philippe Pons, Naomi Russo.

Nicolas Perreaux et Coraline Rey, ont inscrit leurs recherches en master puis en doctorat dans le cadre des CBMA et sont devenus des piliers du programme<sup>10</sup>.

Grâce à ce travail d'équipe, auquel il faut également associer les personnels administratifs et techniques, parallèlement aux projets collectifs, les CBMA ont pu constituer leur propre plateforme documentaire et développer leurs propres projets avec le soutien financier de la Région Bourgogne, du CNRS et de l'université de Bourgogne. Ces apports ont été décisifs, car à l'intérieur d'un laboratoire transpériode et pluridisciplinaire à dominante archéologique, le projet n'a jamais bénéficié de crédits récurrents de l'unité, plutôt polarisés par les chantiers de fouille. Cette tension s'est manifestée notamment lors du tarissement progressif des sources régionales de financement complémentaire à partir de 2011 et du changement de la politique interne qui, à l'inverse des années précédentes, devenait défavorable à notre projet. Le changement de ma propre affectation et l'accueil du programme des CBMA par une équipe d'historiens médiévistes à partir de 2015, le Lamop<sup>11</sup>, a permis de retrouver un cadre propice à des nouveaux projets, mais dépendant, du point de vue des ressources humaines, de contrats de très courte durée dans le contexte déplorable de précarisation croissante des emplois dans le domaine de la recherche<sup>12</sup>. Coraline Rey assume depuis, par intermittence, les tâches d'ingénieure d'études et de webmaster des CBMA.

En dehors de toute situation particulière à tel ou tel projet, ce bref rappel permet de dégager trois points significatifs dans la conduite d'un programme « numérique », et qui relèvent, au dernier ressort, des capacités institutionnelles de financement : le rôle incitatif de la politique nationale de la recherche avec des subventions ciblées, comme celles des appels à projet de l'ANR « Corpus et outils de la recherche en SHS » en 2006, 2007 et 2008, puis en 2011 et 2012 ; l'assise des laboratoires pouvant assurer (ou pas) la continuité sur la longue durée ; la centralité des ressources humaines.

## Principes

Le cadre institutionnel général posé, la structure du programme des CBMA découle de ses orientations éthiques et épistémologiques. Prenant acte du fait que les documents médiévaux sont par essence libres de droits et que le projet est financé par l'argent public, tous les acquis du programme sont mis à disposition librement sur Internet presque

---

<sup>10</sup> C. REY, *L'entreprise archivistique de Jean de Cirey, abbé de Cîteaux (1476-1501). Le dossier documentaire de la seigneurie de Villars en Côte-d'Or*, mémoire de Master 2 sous la direction de E. Magnani et M.-J. Gasse-Grandjean, Université de Bourgogne, Dijon, 2009, disponible en ligne sur <[http://documents.cbma-project.eu/divers/publications/Coraline%20REY\\_memoire%20Master%202\\_2009.pdf](http://documents.cbma-project.eu/divers/publications/Coraline%20REY_memoire%20Master%202_2009.pdf)> et sa thèse en cours *Archives et bibliothèque à Cîteaux, de l'original au « recueil » : scripturalité et visualité dans une entreprise d'écriture au Moyen Âge (XII<sup>e</sup> siècle – début du XVI<sup>e</sup> siècle)*, présentation sur <<http://armaria.hypotheses.org/sujet-de-these>> ; N. PERREAUX, *L'eau, l'écrit et la société. Étude statistique sur les champs sémantiques dans les bases de données [CBMA]*, Mémoire de Master 2 sous la direction d'E. Magnani, Université de Bourgogne, Dijon, 2010, résumé disponible en ligne sur <<http://cem.revues.org/12062>> et sa thèse *L'écriture du monde, op. cit.*, présentée dans les articles « L'écriture du monde (I). Les chartes et les édifices comme vecteurs de la dynamique sociale dans l'Europe médiévale (VII<sup>e</sup>-milieu du XIV<sup>e</sup> siècle) », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 19.2 (2015), disponible en ligne sur <<http://cem.revues.org/14264>> ; DOI : 10.4000/cem.14264 et « L'écriture du monde (II). L'écriture comme facteur de régionalisation et de spiritualisation du *mundus* : études lexicales et sémantiques », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 20.1 (2016), disponible en ligne sur <<http://cem.revues.org/14452>> ; DOI : 10.4000/cem.14452 (liens consultés le 30 mai 2017).

<sup>11</sup> UMR 8589, Lamop (Laboratoire de Médiévisitologie occidentale de Paris - CNRS - Université Paris 1 Panthéon-Sorbonne).

<sup>12</sup> Depuis 2015, en plus de l'appui du Lamop, les CBMA ont reçu le soutien du Consortium Cosme (2016 et 2018) et du Labex haStec (2017, 2018 et 2019).

immédiatement. Plus encore, tout est pensé pour faciliter la consultation, l'utilisation et l'appropriation des textes, des données et des métadonnées tout au long du processus de leur traitement. Cela a une incidence importante sur les choix techniques, car les CBMA se caractérisent par la pluralité des formats électroniques de la publication des documents, et par l'attachement aux formats simples à la portée de tous<sup>13</sup>. Ainsi, les différentes éditions de recueils d'actes ou de cartulaires qui ont été saisies manuellement par un prestataire extérieur (Wordpro, à Pondichéry), contiennent un balisage minimal fondé sur des couleurs (page, numéro, date, texte, etc.) et ont été livrées en format .doc. Les fichiers ont été ensuite ramenés au format basique .txt où les différents champs sont distingués par des simples tabulations ou des paragraphes réguliers. Dans les deux cas, les textes et les premières métadonnées issues des éditions sont aisément exportables vers des tableurs et des bases de données. Nous avons choisi d'utiliser comme outil de travail un logiciel propriétaire, mais éprouvé et de coût modique, FileMakerPro (FMP), très souple pour les opérations d'export et import. Le renseignement de chaque unité documentaire (bénéficiaire, date, diocèse, analyse, référence de l'édition, etc.) se fait sur cette base qui, depuis 2015, est régulièrement convertie en format .csv et publiée. Tous ces formats sont téléchargeables sur le site des CBMA, de même que les versions .pdf des éditions quand elles existent. Il nous semble que l'accès aux formats électroniques de base est une « bonne pratique », indispensable et aussi importante, peut-être davantage que le libre accès aux encodages dont les objectifs sont toujours particuliers, laissant peu de marge de manœuvre aux réappropriations et aux transformations.

Les CBMA ne sont pas pour autant réfractaires au balisage, et une bonne partie des chartes a été encodée en XML en vue de leur interrogation par un logiciel de fouille de texte, *Philologic*<sup>14</sup>. Sur ce point se situe l'engagement épistémologique de l'équipe. En effet, il nous est très vite apparu que la mutation de la recherche historique engendrée par la numérisation en masse de documents médiévaux ne se situait pas dans les bases de données traditionnelles à l'interrogation de type indiciaire, mais dans les méthodes d'analyses lexicales structurales<sup>15</sup>. L'impulsion d'Alain Guerreau et la participation de l'équipe des CBMA à l'ANR OMNIA (Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins - 2008-2013)<sup>16</sup> ont été déterminantes pour les orientations scientifiques du projet. Dans cette perspective, au printemps 2017, ce sont plus de 27 000 chartes déjà enregistrées ainsi que leurs métadonnées, qui ont été harmonisées, lemmatisées et converties dans le format propre à

---

<sup>13</sup> M.-J. GASSE-GRANDJEAN, « Les *Chartae Burgundiae Medii Aevi* (CBMA) et le numérique », in *L'apport des technologies numériques à la diplomatie médiévale, actes de la table ronde organisée par l'Institut historique allemand (R. Große) et l'École nationale des chartes (O. Canteaut), 4<sup>e</sup> journée des sciences humaines et sociales de l'IHA et 7<sup>e</sup> rencontre de la Gallia Pontificia, 20 mai 2011*, in *Francia*, 40 (2013), p. 255-263 ; E. MAGNANI, M.-J. GASSE-GRANDJEAN, N. PERREAUX, et C. REY, « *Chartae Burgundiae Medii Aevi* (CBMA). Du parchemin à l'écran », in *La Lettre de l'INSHS*, 22, mars 2013, p. 27-30, disponible en ligne sur <[http://www.cnrs.fr/inshs/Lettres-information-INSHS/lettre\\_infoinshs\\_22.pdf](http://www.cnrs.fr/inshs/Lettres-information-INSHS/lettre_infoinshs_22.pdf)> (consulté le 30 mai 2017).

<sup>14</sup> <<https://www.lib.uchicago.edu/efts/ARTFL/philologic/>> (consulté le 30 mai 2017).

<sup>15</sup> A. GUERREAU, « Textes anciens en série », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA, Collection CBMA, Les outils*, disponible en ligne sur <<http://cem.revues.org/12177>> ; DOI : 10.4000/cem.12177 (consulté le 30 mai 2017).

<sup>16</sup> B. BON, « OMNIA – Outils et Méthodes Numériques pour l'Interrogation et l'Analyse des textes médiolatins », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 13 (2009), p. 291-292, disponible en ligne sur <<http://cem.revues.org/11086>> ; DOI : 10.4000/cem.11086 ; ID., « OMNIA : outils et méthodes numériques pour l'interrogation et l'analyse des textes médiolatins (2) », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 14 (2010), p. 251-252, disponible en ligne sur <<http://cem.revues.org/11566>> ; DOI : 10.4000/cem.11566 ; ID., « OMNIA : outils et méthodes numériques pour l'interrogation et l'analyse des textes médiolatins (3) », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, 15, 2011, disponible en ligne sur <<http://cem.revues.org/12015>> ; DOI : 10.4000/cem.12015 (consultés le 30 mai 2017).

être importé par le logiciel TXM (Textométrie)<sup>17</sup>, grâce aux algorithmes écrits par Nicolas Perreaux, qui avait déjà œuvré de la sorte pour leur conversion sur *Philologic* en 2010.

La simplicité a été aussi le principe directeur d'un volet supplémentaire de la plateforme des CBMA, la publication, à partir de 2009, de la reproduction en mode image de cartulaires, en collaboration avec les Archives départementales de la Côte d'Or à Dijon<sup>18</sup>. Présentement, plus de 40 manuscrits sont disponibles en format FlipBook et des nouvelles mises en ligne sont en cours de préparation. Ce format a été retenu alors en raison de sa flexibilité d'adaptation, de sa compatibilité avec des nombreux navigateurs et systèmes d'exploitation ainsi que pour son ergonomie de consultation, incluant une option de téléchargement au format .pdf<sup>19</sup>. L'intégration des manuscrits dans la plateforme des CBMA répond à la nécessité d'une approche globale de la production documentaire, y compris dans ses aspects graphiques et visuels. Les textes, en effet, ne sont pas des structures désincarnées, mais constitutifs d'artefacts sensibles qui ont été produits, compris, manipulés et utilisés en tant que tels<sup>20</sup>.

L'équipe des CBMA est bien consciente que son travail, en donnant des nouveaux supports et des possibilités différentes de « lecture »<sup>21</sup>, ajoute d'autres opérations dans la série d'opérations que les documents subissent depuis leur création<sup>22</sup>. Cela nous oblige à ne pas perdre de vue l'incidence de notre propre pratique de re-production et d'enregistrement dans la longue chaîne de transformations et de réactualisations des documents que les procédés numériques ne cessent d'intensifier.

## Mise en ligne et interrogation

La publication sur Internet est la principale différence entre nos projets informatiques actuels et ceux de nos devanciers du dernier tiers du XX<sup>e</sup> siècle. L'expérience des CBMA dans ce domaine est probablement assez exemplaire des cheminements empruntés par d'autres projets numériques, et il me semble utile de revenir rétrospectivement sur ce processus afin de mettre certaines des difficultés rencontrées en perspective.

C'est seulement depuis l'été 2014 que les CBMA ont un site propre, hébergé par la TGIR Huma-Num du CNRS. En 2006, les CBMA étaient simplement signalées par une page

---

<sup>17</sup> <<http://textometrie.ens-lyon.fr/>> (consulté le 30 mai 2017). Le corpus lemmatisé et formaté est téléchargeable ici : <<http://www.cbma-project.eu/bdds2/la-base-sous-txm.html>> (consulté le 22 février 2018).

<sup>18</sup> Dirigées par Gérard Moyse jusqu'en 2013, et ensuite par Édouard Bouyé. Les photographies sont réalisées par Frédéric Petot. Les flipbooks ont été mis en place par M.-J. Gasse-Grandjean et N. Perreaux, et publiés également par C. Rey et A. Dumontet, ingénieur d'études CNRS à l'UMR Artheis.

<sup>19</sup> Depuis, pour profiter des évolutions techniques les plus récentes en matière de mise en ligne et de visualisation de manuscrits, le basculement vers le système iiif (*International Image Interoperability Framework*) est actuellement testé en vue de son implémentation au cours de l'année 2019.

<sup>20</sup> P. RÜCK, « Die Urkunde als Kunstwerk », in *Kaiserin Theophanu. Begegnung des Ostens und Westens um die Wende des ersten Jahrtausends. Gedenkschrift des Kölner Schnütgen-Museums zum 1000. Todesjahr der Kaiserin*, V. A. VON EUW et P. SHREINER (éd.), t. II, Cologne, 1991, p. 311-333 ; *Graphische Symbole in mittelalterlichen Urkunden. Beiträge zur diplomatischen Semiotik*, P. RÜCK (éd.), Sigmaringen, 1996.

<sup>21</sup> F. MORETTI, *Graphes, cartes et arbres. Modèles abstraits pour une autre histoire de la littérature*, trad. E. Dobenesque, Paris, 2008 (*Graphs, maps, trees : abstract models for a literary history*, Londres-New York, 2005).

<sup>22</sup> L. KUCHENBUCH, « Source ou document ? Contribution à l'histoire d'une évidence méthodologique », in *Hypothèses 2003. Travaux de l'École doctorale d'histoire de l'Université Paris I Panthéon-Sorbonne*, Paris, 2004, p. 287-315.

de présentation à l'intérieur du site internet du laboratoire<sup>23</sup>, mais la question de la mise à disposition de tous les acquis du projet s'était posée très tôt<sup>24</sup>. À partir de septembre 2007 et jusqu'en juin 2014, toujours à l'intérieur de ce site collectif dont la configuration, l'hébergement et donc les adresses URL ont changé au moins trois fois au cours de cette période, les CBMA ont déployé une plateforme documentaire qui n'a pas cessé de s'étendre et d'évoluer<sup>25</sup>. En plus des fichiers .doc et .txt, la base de données sur FMP était aussi proposée en forme de tableau HTML<sup>26</sup>, mais ce dernier format était devenu ingérable avec les milliers de chartes engrangées et n'était plus disponible à partir de 2009. Comme l'accès à la base FMP était soumis à une demande d'autorisation auprès de l'équipe, cela signifiait que les métadonnées n'étaient pas joignables directement. Ce verrou est tombé en 2015, au même moment où l'export de la base en .csv vint en quelque sorte remplacer l'ancien tableau HTML de 2007. La version lemmatisée des CBMA sous TXM, publiée en juin 2017, a réglé encore plus fortement le problème, car les métadonnées sont entièrement accessibles et manipulables sous ce format<sup>27</sup>.

Parallèlement à ces mises en ligne, l'équipe s'est engagée dans deux autres développements en vue d'une interrogation experte des textes : d'une part l'interrogation sur *Philologic*, et d'autre part un site pour les CBMA dédié aussi à l'édition électronique, comprenant un système d'interrogation *ad hoc* de la base des données<sup>28</sup>. Disons d'emblée que si le premier projet, malgré un parcours parfois caillouteux, est abouti, le deuxième est resté inachevé.

La volonté de proposer l'interrogation des chartes avec un logiciel de fouille de textes comme *Philologic* s'écartait des standards habituels des bases des données historiennes, et les CBMA demeurent encore aujourd'hui la seule plateforme de documents diplomatiques à offrir ce type d'interrogation. Dès 2006, à l'initiative d'Alain Guerreau, à partir des fichiers .doc et de leur balisage couleur, une version expérimentale des CBMA était mise en ligne par le concepteur du logiciel, Mark Olsen, sur les serveurs de l'université de Chicago. Il fallait toutefois convertir ces essais en une application prenant en compte les enrichissements en métadonnées des CBMA, ce qui nous a conduit à formuler en 2008 le projet d'installer la base et le logiciel d'analyse sur un serveur propre. Une fois les crédits demandés et obtenus auprès de l'université de Bourgogne, via la MSH (préciput ANR), le matériel a été acquis en 2009. Toutefois il a été impossible de trouver de la place dans un local adéquat pour l'installer (en fait, réfrigéré et pourvu d'une connexion Internet) dans toute l'université de Bourgogne, qui, bien entendu, en est pourvue. Aucun des services informatiques de l'université n'a voulu accueillir ce serveur, ni le service général, ni ceux de l'UFR Sciences de la Terre qui loge le

<sup>23</sup> <<https://web-beta.archive.org/web/20061116211111/http://www.archeologie-cultures-societes.cnrs.fr/recherche/CBMA.htm>> (capture du 16 novembre 2006, consulté le 30 mai 2017).

<sup>24</sup> Le parcours que je reformule ici peut être reconstitué en grande partie par les comptes rendus des journées d'études des CBMA réunies annuellement entre 2007 et 2014, disponibles en ligne sur <<http://cem.revues.org/12172>> (consulté le 30 mai 2017).

<sup>25</sup> <<https://web-beta.archive.org/web/20070521052037/http://www.archeologie-cultures-societes.cnrs.fr:80/recherche/CBMA.htm>> (capture du 21 mai 2007, consulté le 30 mai 2017).

<sup>26</sup> <<https://web-beta.archive.org/web/20080917201531/http://www.artehis.cnrs.fr:80/BDD/CBMA/CBMA%20to%20web%20tri%20rubriques.htm>> (capture du 17 septembre 2008, consulté le 30 mai 2017).

<sup>27</sup> Actuellement (février 2019), l'équipe étudie aussi l'évolution de la base de travail FMP vers SQLite et son interrogation en ligne avec CQP-Web.

<sup>28</sup> Au même moment, en vue de la diffusion et de la valorisation du programme, les CBMA se sont appuyés sur une autre plateforme collective, OpenEdition (<<http://www.openedition.org/>>), avec une collection publiant des articles de différentes ordres sur Revues.org, au sein du site de la revue BUCEMA (<<http://cem.revues.org/12172>>) et un blog sur Hypothèses.org (<<http://laetusdiaconus.hypotheses.org/>>) (liens consultés le 30 mai 2017).

laboratoire, ni ceux de la MSH de Dijon. Le serveur est finalement installé dans une salle du laboratoire, grâce à l'opiniâtreté de l'équipe et les compétences informatiques d'un doctorant<sup>29</sup>. À l'été 2010, le serveur marche parfaitement et une journée d'études se tient début 2012 pour discuter des recherches effectuées avec cet outil<sup>30</sup>. Mais la salle en question étant réquisitionnée pour d'autres usages sans autre proposition de relocalisation convenable nous oblige à abandonner le serveur propre et à envisager une solution externe. Et en effet, à partir de septembre 2012 l'application *Philologic* des CBMA est hébergée sur les serveurs du Très Grand Équipement ADONIS à l'IN2P3 de Lyon, avec l'aide à l'installation de Gérard Foliot. Cet hébergement providentiel perdure dans le cadre de la refonte du TGE ADONIS à l'intérieur du TGIR Huma-Num en 2013. Il décharge l'équipe d'une série de contraintes techniques et de manutention, mais il a l'inconvénient de l'écarter de l'administration du serveur. Une conséquence fâcheuse de cet éloignement géographique est qu'il a depuis été techniquement impossible de mettre à jour l'application avec les nouvelles incorporations de chartes<sup>31</sup>. Quoiqu'il en soit, les 13 379 actes interrogeables avec *Philologic*, dont les 5 000 chartes de l'abbaye de Cluny, permettent déjà des recherches fécondes, en attendant la publication du corpus complet sous *Philologic4* en 2019.

Le deuxième projet, qu'on avait intitulé provisoirement « CBMA on-line », comptait élaborer une application *ad hoc* pour convertir pour le web la base de données de travail sur FMP et proposer une interrogation avancée, le tout sur un site propre pour les CBMA. Ce projet nous situe dans le contexte de l'institutionnalisation des Humanités Numériques, telle qu'elle se met place dans le domaine des sciences humaines et sociales, à l'initiative du Département devenu Institut des sciences humaines et sociales du CNRS. Il convient de le rappeler brièvement.

À juste titre, le CNRS a compris les besoins et le rôle que des infrastructures collectives devaient jouer pour accompagner et développer les recherches dans leur volet numérique. En 2005, une UPS (Unité Propre de Service n° 2916) intitulée « Accès unique aux documents numériques en sciences humaines et sociales » (ADONIS) est créée. Sous l'égide de ce qui devient en 2007 le Très Grand Équipement ADONIS (Accès unifié aux Données et documents Numériques des Sciences humaines et sociales), commencent à être créés également, à partir de 2005, des Centres de ressources numériques (CRN), adossés à des laboratoires ayant déjà un savoir-faire dans différents types d'objets (écrits, sonores, visuels, SIG, 3D ...). En 2010, il existait une dizaine de CRN<sup>32</sup>. Dans le domaine de l'histoire médiévale le CRN référent est TELMA (Traitement électronique des manuscrits et des archives), porté entre 2006 et 2010 conjointement par l'École nationale des Chartes (EnC) et l'Institut de Recherche et d'Histoire des Textes (IRHT, unité de recherche propre du CNRS), et qui depuis 2010 est seul à abriter ce CRN<sup>33</sup>. Pour compléter ce tableau, notons la création en 2011 de la TGIR (Très grande infrastructure de recherche) Corpus-IR pour la mise en place de consortiums, qui a fusionné en 2013 avec le TGE ADONIS pour devenir la TGIR

<sup>29</sup> Nicolas Perreaux a adapté le programme, avec l'aide de Mark Olsen, Alain Guerreau et le suivi de Marie-José Gasse-Grandjean. Il s'est occupé également de l'installation de l'application dans les serveurs du TGE ADONIS.

<sup>30</sup> Compte-rendu dans E. MAGNANI et M.-J. GASSE-GRANDJEAN, « CBMA. *Chartae Burgundiae Medii Aevi*. VI. Les chartes bourguignonnes sous *PhiloLogic* », *Bulletin du centre d'études médiévales d'Auxerre | BUCEMA*, Collection CBMA, Les journées d'études, disponible en ligne sur <<http://cem.revues.org/12595>> (consulté le 30 mai 2017).

<sup>31</sup> Depuis la rédaction de cet article, cette difficulté a été contournée grâce au travail de Pierre Brochard, ingénieur d'études au CNRS, rattaché au Lamop depuis juin 2017. Une version à jour des CBMA sous *Philologic4* paraîtra courant 2019.

<sup>32</sup> Voir le diaporama de S. POUYLLAU et R. WALTER, *La politique du TGE pour les centres de ressources dans le cadre des infrastructures nationales*, disponible en ligne sur <[http://www.huma-num.fr/sites/default/files/valpre\\_crn.pdf](http://www.huma-num.fr/sites/default/files/valpre_crn.pdf)> (consulté le 30 mai 2017).

<sup>33</sup> <<http://www.cn-telma.fr/a-propos>> (consulté le 30 mai 2017).



Huma-Num<sup>34</sup>. Dans toutes ces initiatives, on perçoit l'objectif d'œuvrer à la fédération des ressources humaines et techniques.

Dans ce contexte, porteurs d'une application dédiée aux documents diplomatiques, c'est tout naturellement que nous nous sommes tournés en 2007 vers l'EnC. Nous avons alors élaboré conjointement un cahier des charges, passé une convention, puis reçu un devis et versé des crédits. Malgré toute la bonne volonté et l'engagement des responsables des services informatiques de l'EnC à l'époque, sans les personnels nécessaires pour mener leurs propres projets, prioritaires par rapport à ceux des extérieurs comme le nôtre, la réalisation du produit attendu pour 2008 puis pour 2009, n'a pas vu le jour. Un mémoire de stage soutenu en septembre 2011 a cependant engagé la réflexion pour la conversion XML/TEI de la base des CBMA, et sur cet acquis une application de conversion a été élaborée<sup>35</sup>, mais on était loin du site et de l'interrogation experte que l'on avait imaginé. En fait, les CRN recevaient des subventions mais n'étaient pas dotés de postes pérennes supplémentaires pour assurer leur mission. On était donc toujours sur le modèle de fonctionnement basé sur des emplois précaires et des contrats de courte durée. Dans ce cadre les avancées ne peuvent se faire qu'au coup par coup.

Ces incidents de parcours, qui peuvent paraître anecdotiques, sont à l'inverse révélateurs d'un mouvement de fond : celui de la création d'un champ académique en voie de structuration, dont on n'est pas toujours conscient de participer à son délinéament, tout comme, pour reprendre les formules de Thomas Kuhn, de vivre une période de « crise » de la « science normale » qui prélude une « révolution scientifique »<sup>36</sup>. Même si les changements semblent s'opérer très vite en vue de la course effrénée de l'informatique, on travaille encore sur des paradigmes éprouvés qui laissent peu de marge aux nouvelles manières de faire. Ce décalage qu'on ressent au niveau institutionnel – on accorde les moyens pour l'acquisition du matériel nécessaire mais pas pour son installation et sa mise en marche ; on constitue des plateformes collectives sans les effectifs humains suffisants –, mais aussi dans la résistance et le scepticisme de nombreux chercheurs, sont autant d'indices d'une pratique qui n'est pas reconnue ni comprise pleinement.

## Corpus

En fait, l'intérêt d'un programme numérique en SHS réside dans son potentiel explicatif ou heuristique. La possibilité d'interroger une masse inédite de documents médiévaux ne peut être véritablement renovatrice pour la recherche que si l'on utilise les méthodes adaptées à cette forme et à cette échelle d'analyse, mais aussi que si l'on accepte le décentrement que ces approches – sémantiques, statistiques, spatiales ... – impliquent par rapport à nos méthodes traditionnelles, et que, comme tout scientifique, l'on se donne le droit à l'expérience. Pour traiter des textes en série on n'est pas dépourvu, en sciences humaines et

---

<sup>34</sup> Cf. le rapport d'activité 2013-2015 de la TGIR Huma-Num disponible en ligne sur <<http://www.huma-num.fr/sites/default/files/Rapport-activite-2013-2015.pdf>> (consulté le 30 mai 2017).

<sup>35</sup> S. DELLA BARTOLOMEA, *Analyse et modélisation XML/TEI pour l'édition électronique des Chartae Burgundiae Medii Aevi*, Mémoire de stage sous la direction de F. Clavaud, Master 2 « Technologies numériques appliquées à l'histoire », École nationale des Chartes, Année 2010-2011.

<sup>36</sup> Th. KUHN, *La structure des révolutions scientifiques*, trad. L. Meyer, Paris, 1983 [*The Structure of Scientific Revolutions*, Chicago, 1962, 2<sup>e</sup> éd. revue et augmentée 1970, trad. franç. 1972]. Voir N. PERREAUX, « De l'accumulation à l'exploitation ? Expériences et propositions pour l'indexation et l'utilisation des bases de données diplomatiques », in *Digital diplomatics. The computer as a tool for the diplomatist?*, A. AMBROSIO, S. BARRET, G. VOGELER (éd.), Köln-Weimar-Wien, 2014 (*Archiv für Diplomatik. Schriftgeschichte Siegel- und Wappenkunde*, Beiheft 14), p. 187-210.

sociales, de précédents théoriques et techniques importants, comme les réflexions, depuis les années 1950, de la linguistique de corpus (*Corpus Linguistic*). De ces recherches émerge l'idée qu'à partir des usages linguistiques attestés, repérés et réunis en corpus, on peut aussi bien vérifier une hypothèse afin de la valider, la réfuter ou l'améliorer (on parle alors d'approche « *corpus-based* »), que l'explorer sans a priori pour faire émerger du corpus lui-même des hypothèses sur le langage (approche dite « *corpus-driven* »)<sup>37</sup>. Ce dernier cas revient à admettre que le sens se situe à l'échelle du corpus lui-même, et pas dans celle du mot, de la phrase ou du texte isolés. C'est dans cette direction, nouvelle pour les historiens médiévistes, que nous pensons que les documents numérisés et réunis « en corpus » par les CBMA peuvent être utiles pour la recherche, en tant que terrain propice aux expériences nouvelles.

Alors qu'avec les actes réunis par les CBMA on a déjà pu montrer que la cohérence de la production documentaire bourguignonne, ses flux chronologiques et spatiaux, correspondaient à des dynamiques sociales différenciées<sup>38</sup>, il nous semble que nous ne pouvons pas nous contenter des seuls textes diplomatiques, en particulier dans une région également prolifique en matière de production de textes narratifs, théologiques et normatifs, propres à resserrer, voire à transformer, les interprétations. Dans cette perspective, nous nous sommes engagés dans la constitution d'un corpus hétérogène et structuré de documents latins médiévaux bourguignons de tous genres entre le V<sup>e</sup> et le XV<sup>e</sup> siècle. Cet engagement se traduit dans le titre même du projet où dans l'acronyme CBMA, les Chartae deviennent Corpus<sup>39</sup>. Depuis l'automne 2016 l'équipe est en train de réunir un sous-corpus de textes hagiographiques, estimés à 380 unités, et d'autres types documentaires suivront à moyen terme<sup>40</sup>. L'objectif est de pouvoir interroger tous les textes ensemble, mais aussi par sous-groupe typologique.

Une première évolution importante dans la base de données a été l'adaptation, par Coraline Rey, des séries de métadonnées qui avaient été pensées pour les chartes pour qu'elles puissent correspondre aussi aux vies de saints, complétés par l'implémentation de nouveaux champs permettant de prendre en compte les particularités de cette typologie documentaire. Le site internet des CBMA a été aussi repensé afin de répondre à ces nouvelles incorporations. Un changement significatif réside dans la granularité plus fine de présentation, et le passage d'un affichage par « édition » à celle par « texte ». En fait, les « éditions » de chartes, auxquelles on accède toujours en tant que telles dans la plateforme, sont le plus souvent des recueils construits par les éditeurs à partir des actes relatifs à une institution productrice/réceptrice (monastère, église cathédrale ...), ce qui est un critère parmi d'autres possibles pour assembler ce qu'on pourrait désigner comme des « proto-corpus ». Le fait qu'aucune édition parmi celles numérisées par les CBMA ne corresponde à la transcription d'un cartulaire médiéval présent dans la plateforme en mode image (FlipBook) est à ce titre évocateur du procédé de « re-fabrication » réalisé par les éditions modernes.

---

<sup>37</sup> Les références bibliographiques étant très nombreuses, nous nous limitons à citer ici E. TOGNINI-BONELLI, *Corpus Linguistics at Work*, Amsterdam, 2001 (*Studies in Corpus Linguistics*, 6) et pour la France, la revue *Corpus* (<<https://corpus.revues.org/>>) et les travaux de D. MAYAFFRE (par exemple, « Les corpus réflexifs : entre architextualité et hypertextualité », *Corpus*, 1, 2002, p. 51-69, disponible en ligne sur <<http://corpus.revues.org/index11.html>> (consulté le 30 mai 2017).

<sup>38</sup> N. PERREAUX, *L'écriture du monde. op. cit.*

<sup>39</sup> E. MAGNANI, « Qu'est-ce qu'un corpus ? », *Les carnets de l'IRHT*, 2 octobre 2017, disponible en ligne sur <<https://irht.hypotheses.org/3187>> (consulté le 10 février 2019).

<sup>40</sup> E. MAGNANI, « Un corpus structuré et hétérogène de textes latins médiévaux (Bourgogne, V<sup>e</sup>-XV<sup>e</sup> siècle) », *Bulletin du CERCOR*, 41 (2017), p. 59-65. Ce projet, en plus du soutien du Lamop, a reçu le soutien du Consortium Cosme (2016 et 2018) et du Labex haStec (2017, 2018 et 2019).

Cette disjonction, dont on ne mesure pas toujours la portée, résulte des opérations de réactualisation subies par les documents dans le temps. En adoptant la notion de « texte », qui ne va pas forcément de soi, nous avons voulu indiquer une unité documentaire, qu'elle soit une charte, une vie de saint ou autre, et que l'on pourra mettre en relation les unes par rapport aux autres en raison de leur rattachement à une institution ou de n'importe quel autre critère, y compris ceux qu'une analyse *corpus-driven* pourra rendre explicites et que nos préconceptions ne sauraient pas envisager. Pour interroger ces documents, les CBMA proposeront les textes lemmatisés dans un format propre à importer le corpus dans l'application de fouille de texte Textométrie (TXM) que chacun peut facilement installer directement dans son propre ordinateur. Les textes seront aussi renseignés avec des données de géolocalisation, en vue de soumettre le corpus à des exploitations spatiales<sup>41</sup>. Avec ces transformations décisives, les CBMA entendent continuer à proposer des ressources raisonnées et structurées, aptes aux réappropriations, pour une recherche ouverte et attentive aux expériences et aux méthodes propres aux analyses de la masse de documents numérisés<sup>42</sup>.

## Résumé

Depuis fin 2004, le programme des *Chartae/Corpus Burgundiae Medii Aevi* s'est consacré à la mise en ligne d'actes diplomatiques bourguignons sous plusieurs formats électroniques mais aussi de cartulaires en mode image à feuilleter. Dès le départ il s'est orienté vers des choix techniques simples et une mise à disposition presque immédiate de tous ses acquis en vue d'une réappropriation aisée par les chercheurs. D'un point de vue structurel, le projet a connu aussi bien les limites que les avancées des plateformes collectives institutionnelles : les services informatiques universitaires, les centres de ressources numériques et les très grands équipements du CNRS. Diffusé pendant plusieurs années à l'intérieur d'un site de laboratoire, il a développé un site internet indépendant en 2014, après avoir abandonné l'initiative d'entretenir un serveur propre. Sur la base d'un solide ensemble diplomatique, le projet s'oriente maintenant vers la constitution d'un corpus hétérogène et structuré incorporant des textes narratifs, exégétiques, entre autres, dans la perspective d'alimenter une recherche, qu'on est en droit d'attendre en renouvellement, grâce à l'exploitation raisonnée de larges corpus.

## Abstract

Since the end of 2004, the program of the *Chartae / Corpus Burgundiae Medii Aevi* has devoted itself to putting on line Burgundian diplomatic acts in several electronic formats but also cartularies to be flipped. From the outset he has been oriented towards simple technical choices and an almost immediate disposition of all his achievements with a view to

---

<sup>41</sup> Sont renseignées, selon les systèmes de projection Lambert 93 et WGS 84 - EPSG 4326, les coordonnées du lieu auquel est rattaché le saint, les coordonnées du lieu de rédaction d'un texte, et les coordonnées de l'institution promotrice. Sur les avancées du corpus hagiographique voir E. MAGNANI, « Les CBMA en corpus structuré. Atelier 2. Le corpus hagiographique bourguignon. Débats et recherches », *Bulletin du centre d'études médiévales d'Auxerre, BUCEMA*, Collection CBMA, Les journées d'études, disponible en ligne dans < <http://journals.openedition.org/cem/15493> > (consulté le 10 février 2019). Par ailleurs, en 2018, en collaboration avec le CESCO de Poitiers, l'équipe a mis en place un corpus de plus de 1400 inscriptions épigraphiques (cf. E. MAGNANI, E. INGRAND-VARENNE, « Le corpus épigraphique bourguignon (VIII<sup>e</sup>-XV<sup>e</sup> siècle). Des catalogues aux applications numériques », *Bulletin du centre d'études médiévales d'Auxerre, BUCEMA*, Collection CBMA, Les journées d'études, disponible en ligne dans < <http://journals.openedition.org/cem/15591> > (consulté le 10 février 2019).

<sup>42</sup> Je remercie Nicolas Perreaux, Coraline Rey et Daniel Russo pour leurs relectures de cet article et leurs suggestions d'améliorations.

an easy reappropriation by the researchers. From a structural point of view, the project has encountered both the limitations and the progress of institutional collective platforms: university IT services, digital resource centers and very large CNRS equipment. Diffused for several years inside a laboratory site, it developed an independent website in 2014, after having abandoned the initiative to maintain its own server. On the basis of a solid diplomatic ensemble, the project is now oriented towards the constitution of a heterogeneous and structured *corpus* incorporating narrative and exegetical texts, among others, with a view to feeding a research, right to expect renewal, thanks to the reasoned exploitation of large *corpora*.

Mots-clés :

corpus, histoire, diplomatique, hagiographie, Humanités Numériques, Moyen Âge

Keywords :

corpus, history, diplomatic, hagiography, Digital Humanities, Middle Ages