



**HAL**  
open science

## Exon Domesday : méthodes numériques appliquées à la codicologie pour l'étude d'un manuscrit anglo-normand

Peter Anthony Stokes, Geoffroy Noël

### ► To cite this version:

Peter Anthony Stokes, Geoffroy Noël. Exon Domesday : méthodes numériques appliquées à la codicologie pour l'étude d'un manuscrit anglo-normand. *Tabularia : Sources écrites des mondes normands médiévaux*, 2019, 10.4000/tabularia.4118 . hal-02441569

**HAL Id: hal-02441569**

**<https://ephe.hal.science/hal-02441569v1>**

Submitted on 2 Oct 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.

## Exon Domesday : méthodes numériques appliquées à la codicologie pour l'étude d'un manuscrit anglo-normand

*Exon Domesday: Digital Methods in Codicology for the Study of an Anglo-Norman Manuscript*

*Exon Domesday: Metodi digitali per lo studio codicologico di un manoscritto anglo-normanno*

**Peter A. Stokes et Geoffroy Noël**

---



**Édition électronique**

URL : <http://journals.openedition.org/tabularia/4118>

DOI : 10.4000/tabularia.4118

ISSN : 1630-7364

**Éditeur :**

CRAHAM - Centre Michel de Boüard, Presses universitaires de Caen

**Référence électronique**

Peter A. Stokes et Geoffroy Noël, « Exon Domesday : méthodes numériques appliquées à la codicologie pour l'étude d'un manuscrit anglo-normand », *Tabularia* [En ligne], Les sources des mondes normands à l'heure du numérique, mis en ligne le 20 décembre 2019, consulté le 10 janvier 2020.

URL : <http://journals.openedition.org/tabularia/4118> ; DOI : 10.4000/tabularia.4118

---



# Exon Domesday : méthodes numériques appliquées à la codicologie pour l'étude d'un manuscrit anglo-normand

## *Exon Domesday: Digital Methods in Codicology for the Study of an Anglo-Norman Manuscript*

## *Exon Domesday: Metodi digitali per lo studio codicologico di un manoscritto anglo-normanno*

Peter A. STOKES

École Pratique des Hautes Études – Université PSL, Paris  
peter.stokes@ephe.psl.eu

Geoffroy NOËL

King's College London, King's Digital Laboratory, London  
geoffroy.noel@kcl.ac.uk

### Résumé :

Dans cet article, nous présentons une analyse méthodologique de l'étude de l'Exon Domesday Book, Exeter, Cathedral Library, ms 3500. L'Exon Domesday contient la documentation relative au sud-ouest de l'Angleterre, compilée au cours de l'enquête Domesday, puis retravaillée dans le *Great Domesday Book*, tel que nous le connaissons aujourd'hui. L'Exon Domesday offre donc un point de vue unique sur les méthodes de travail des premiers anglo-normands, nous fournissant des témoignages essentiels de ce qu'ils ont fait, comment ils l'ont fait, et probablement aussi pourquoi. Cependant, le manuscrit est extrêmement complexe, car il est le fruit du travail d'une équipe de scribes qui travaillent ensemble dans des délais très serrés. Sa structure est très irrégulière, étant formé de cahiers allant d'un bifeuillet, ou même d'un feuillet dépareillé, jusqu'à un cahier de vingt feuillets. Afin de comprendre le manuscrit et sa fabrication, nous avons donc développé des outils et des méthodes numériques et pour aider à visualiser et analyser sa structure. Bien qu'ils soient fermement ancrés dans les problématiques du projet, ces outils sont également fondés sur des principes de codicologie qui, à notre avis, peuvent s'appliquer beaucoup plus largement au-delà de ce projet particulier.

Mots clés : Domesday Book, codicologie, humanités numériques, Angleterre anglo-normande

### Abstract:

*In this article, we present a methodological discussion of a study of the Exon Domesday Book, Exeter, Cathedral Library, ms 3500. Exon Domesday contains records for the south-west of England which were compiled for the Domesday survey and then reworked into Great Domesday Book as we know it today. Exon Domesday therefore provides a unique view into the workings of the early Anglo-Normans, providing us with crucial evidence of what they did,*

*how and probably also why. However, the manuscript is extremely complex, being the work of a team of scribes working together under significant pressures of time. It is very irregular in composition, with quires ranging from one bifolium or even a singleton to those of twenty leaves. In order to understand the manuscript and its production, we have developed digital tools and approaches to help visualise and analyse the structure. Although they are based firmly in the research questions of the project, they are also founded on principles of codicology that we believe are applicable much more widely than just this project.*

*Keywords : Domesday Book, codicology, digital humanities, Anglo-Norman England*

*Riassunto :*

*In questo articolo presentiamo una discussione metodologica relativa allo studio del manoscritto Exon Domesday Book, Exeter, Cathedral Library, ms 3500. L'Exon Domesday contiene documenti relativi al sud-ovest dell'Inghilterra compilati nel corso dell'inchiesta relativa alla compilazione del Domesday Book e poi rielaborati nel Great Domesday Book come lo conosciamo oggi. L'Exon Domesday fornisce quindi una visione unica del funzionamento dei primi anni di dominio anglo-normanno, fornendoci le prove cruciali di ciò che hanno fatto, ma anche di come lo hanno fatto e probabilmente del perché. Tuttavia, il manoscritto è estremamente complesso, trattandosi del lavoro di un gruppo di copisti che lavoravano insieme con tempi di realizzazione molto stretti. La struttura del manoscritto è molto irregolare, con fascicoli che vanno dal bifolio o anche composti di una carta isolata, fino a fascicoli di venti fogli. Per comprendere il manoscritto e la sua fabbricazione, sono stati sviluppati strumenti e approcci digitali che aiutano a visualizzarne e analizzarne la struttura. Anche se tali strumenti sono stati concepiti per rispondere alle questioni di ricerca specifiche del progetto, questi si basano anche su principi di codicologia che riteniamo applicabili ampiamente.*

*Parole chiave : Domesday Book, codicologia, informatica umanistica, Inghilterra anglo-normanna*

L'importance du Great Domesday Book (ci-après GDB) pour les études anglo-normandes est bien connue<sup>1</sup>. Compilé par les commissaires de Guillaume I<sup>er</sup> en 1086, son but semble clairement avoir été d'enregistrer tous les biens imposables en Angleterre. Il fournit ainsi des informations extraordinairement détaillées et étendues sur l'Angleterre rurale, sa structure sociale, son paysage et ses ressources. Cependant, le GDB lui-même est en fait un résumé d'autres documents plus détaillés qui ont été compilés dans le cadre de la même enquête, dont le seul survivant presque en totalité pourrait être le manuscrit Exeter, Cathedral Library,

---

1. Great Domesday Book : Londres, TNA, E 21/21. La plupart de ces paragraphes d'introduction sont résumés ou paraphrasés à partir de la proposition de financement du projet qui a été écrite principalement par Julia Crick avec Stephen Baxter et Peter Stokes. L'article dans son ensemble s'appuie énormément sur le travail de toute l'équipe de projet (voir le site Web du projet) et est le résultat d'un financement du Arts and Humanities Research Council (AHRC) no. AH/L013975/1. Sauf indication contraire dans la section concernée, tous les développements logiciels ont été réalisés par Geoffroy Noël, et l'analyse (au sens logiciel d'examen des besoins, de visualisation des données, et ainsi de suite) a été réalisée par Geoffroy Noël et Peter Stokes. Nous remercions l'équipe du projet et plus spécialement Frank Thorn pour la traduction du texte d'Exon Domesday Book et pour ses conseils sur des termes anglo-normands spécialisés.

ms 3500, mieux connu sous le nom d'Exon Domesday Book (ci-après EDB)<sup>2</sup>. À l'heure actuelle, l'EDB comprend un peu plus de 530 feuillets pour plus de 100 cahiers. Ce manuscrit contient des documents concernant le sud-ouest de l'Angleterre, plus précisément les comtés de Cornwall, Devon, Dorset, Somerset et Wiltshire (bien que tous les comtés ne soient pas entièrement documentés dans le manuscrit, probablement parce que le manuscrit a perdu des feuillets et sans doute des cahiers au fil des siècles). Le contenu se compose de différents types de documents, le plus fréquent étant celui des descriptions de fiefs classées d'abord par comté, puis par tenant en chef.<sup>3</sup> Un exemple d'entrée est le suivant<sup>4</sup> :

*Regina Mathildis habuit .i. mansionem que uocatur Litelfroma quam tenuit Bristricus ea die qua rex Eduuardus fuit uiuus et mortuus et reddidit gildum pro .xiii. hidis. Has possunt arare .viii. carruce, et modo tenet Willelmus rex in dominio. De his .xiii. hidis tenet rex .x. hidas et dimidiam et .iii. carrucas in dominio et villani .ii. hidas et dimidiam et .iii. carrucas. Ibi habet rex .x. uillanos et .iii. bordarios et .viii. cotarios et .vi. seruos et .ii. roncinos et .xix. animalia et .ccc. oues et .l. capras et .i. molendinum qui reddit .iiii. solidos et .viii. quadragenarias nemoris in longitudine et .vi. in latitudine et .x. agros prati et .xx. quadragenarias pascuae in longitudine et .ii. in latitudine. Haec mansio ualet per annum .xviii. libras et quando Aiulfus recepit ualebat .xii. libras.*

La reine Mathilde avait un domaine rural qui s'appelle Litelfroma [Little Frome, Dorset] que Beorhtric a tenu le jour où le roi Édouard vivait et était mort, et il rapportait le *geld* pour 13 *hides*<sup>5</sup>. 8 charrues peuvent labourer celles-ci et maintenant le roi Guillaume tient ceci en propriété domaniale. Entre ces 13 *hides* le roi tient 10 *hides* et demi et 3 charrues dans sa réserve et les vilains [tiennent] 2 *hides* et demi et 3 charrues. Là le roi a 10 vilains et 3 bordiers et 8 cotiers<sup>6</sup> et 6 serfs et 2 roussins et 19 bovins et 400 moutons et 50 chèvres et un moulin, qui rapporte 4 sous; et 8 quarantaines de bois en longueur et 6 en largeur et 10 acres de pré et 20 quarantaines de pâturages en longueur et en largeur. Ce domaine rural vaut 18 livres par an et, quand Aiulf l'a reçu, il valait 12 livres.

Comme le montre cet exemple, le texte donne beaucoup de détails, à un niveau qui n'a pas été retenu pour le GDB, et dont nous ne bénéficions donc que pour le sud-ouest de l'Angleterre; son importance a depuis longtemps été

- 
2. La place du Little Domesday Book (LDB : Londres, TNA E 31/1/3) dans la généalogie de GDB est incertaine, mais il pourrait représenter la même phase qu'EDB. Voir, par ex., CLARKE 1985, p. 50-70, en particulier la tabulation sur la page 52.
  3. Tenant en chef : *tennants in chief*, les barons qui tiennent leurs terres directement du roi.
  4. THORN 2018, « Entry » 29a2, <http://www.exonduomesday.ac.uk/digipal/manuscripts/1/texts/view/?above=location/entry/29a2>.
  5. *Geld* : impôt foncier parfois annuel calculé sur la base des *hides*. Une *hide* est une estimation très ancienne de l'étendue imposable d'un terrain, surtout utilisée comme terme pour les terres arables. Elle n'était pas une mesure fixe, mais une unité fiscale et de services dus soit au seigneur soit au comté soit au roi. Nous remercions Frank Thorn pour la traduction du texte et pour ces précisions.
  6. Le « cotier », du latin *cotarius*, est l'habitant d'une petite maison (de campagne).

reconnue par les historiens, et notamment sa valeur pour la compréhension de l'histoire économique et sociale de l'Angleterre du XI<sup>e</sup> siècle. Il est aussi manifestement d'un grand intérêt pour les historiens locaux et même pour les résidents des régions qu'il décrit, particulièrement le doyen et le chapitre de la cathédrale d'Exeter, ainsi que les « Friends » qui ont formé « une organisation caritative pour coopérer avec le doyen et le chapitre afin d'aider à préserver et à améliorer le tissu et l'ameublement » de la cathédrale<sup>7</sup>, et pour lesquels sont particulièrement importantes la diffusion et la valorisation d'un manuscrit qui leur appartient.

En dépit de cette importance de l'EDB en tant que témoin critique de l'enquête Domesday et de la production du GDB, on a accordé beaucoup moins d'attention scientifique à celui-ci par rapport à celle dont a bénéficié le GDB, particulièrement en ce qui concerne le manuscrit lui-même. De plus, des questions fondamentales demeurent sur l'EDB et sa composition. Le matériel y est présenté dans l'ordre féodal, c'est-à-dire d'abord par fief puis par comté, ce qui contraste avec le GDB qui est présenté dans l'ordre territorial, à savoir par comté. Comment le livre a-t-il donc été compilé et quelles sources ont été utilisées ? Comment les nombreux scribes ont-ils collaboré et d'où venaient-ils ? Le matériel dans l'EDB reflète une organisation sous-jacente par *hundred*, qui était la circonscription de base dans le système juridique anglais avant la conquête<sup>8</sup> ; quand et comment ce matériel a-t-il été réorganisé dans le cadre féodal ? Était-ce pendant les séances de l'enquête elle-même ou dans le cadre d'une étape distincte ? Comment ce processus de compilation se rapporte-t-il à celui du GDB ? Même l'ordre original des feuillets et des cahiers est inconnu. Le manuscrit contient plusieurs foliotations différentes, indiquant qu'il a été réordonné et relié plusieurs fois, mais l'ordre original est inconnu, et il n'est même nullement évident que les pages étaient reliées au moment de sa production originale et de ses premières utilisations<sup>9</sup>.

C'est une tentative de réponse à ces questions qui a mené au projet Exon Domesday. Exécuté de 2014 à 2018 et basé conjointement au King's College de Londres et à l'Université d'Oxford, ce projet comportait une étude paléographique et codicologique détaillée du manuscrit, y compris de son texte, ainsi que du contexte historique de sa production, de ses premières utilisations et de sa relation avec le GDB. Il comprend la création d'un site web présentant des images numériques de haute qualité avec un texte latin entièrement nouveau, une traduction en anglais, une description paléographique et codicologique, et d'autres données et visualisations, présentées dans une version adaptée du logiciel Archetype qui donne des moyens d'analyser et de présenter le contenu

---

7. Pour les *Friends of Exeter Cathedral*, voir <https://www.exeter-cathedral.org.uk/support-us/the-friends-of-exeter-cathedral/>.

8. THORN et THORN 1979-1985.

9. Pour la discussion sur la commande et ses changements voir aussi KER, 2002, vol. 2, p. 800-801. Le vocabulaire codicologique utilisé ici suit MUZERELLE, 2002-2003, spécialement III.1.1, « Feuilles et feuillets ».

paléographique et textuel des documents écrits<sup>10</sup>. Cet article présentera donc certaines des questions scientifiques et méthodologiques du projet et discutera de la manière dont elles ont été abordées : on s'intéressera en particulier à la manière dont nous avons modélisé numériquement la structure codicologique du manuscrit pour apporter de nouveaux éléments d'analyse relatifs aux questions posées par le projet.

## La codicologie

La structure du manuscrit de l'EDB est extrêmement complexe, au point qu'il est difficile même pour les experts de la comprendre, et encore plus pour un public plus large. Cependant, cette structure et sa complexité même fournissent des détails cruciaux sur l'enquête du Domesday et de la manière dont elle a été menée sur le terrain. Par exemple, si le manuscrit avait été un simple exemplaire créé par une seule personne à un moment donné, on s'attendrait à ce qu'il soit construit en cahiers réguliers de huit feuillets chacun, à la taille et à la réglure cohérentes. Au contraire, si la taille de page typique de l'EDB est d'environ 260 x 165 mm, avec un cadre de réglure généralement autour de 195 x 135, la plupart des pages ayant vingt lignes, il existe une variation considérable parmi tous ces chiffres. La hauteur des pages varie entre 175-266 mm et la largeur entre 127-179 mm. Le cadre de réglure varie entre 92-212 mm de hauteur et 92-150 mm de largeur. Le nombre de lignes tracées par page varie de 9 à 40, avec quarante lignes trouvées de manière cohérente et exclusive dans ce qui est aujourd'hui le premier cahier, mais qui est irrégulier ailleurs dans le manuscrit. En ce qui concerne la piqûre, quatre outils différents semblent avoir été utilisés : 242 feuillets ont des piqûres en forme de Y, 196 ont des fentes, 167 ont des piqûres en forme de X et un a des doubles perforations. Bien que le nombre le plus fréquent de feuillets par cahier soit de huit, ce qui était normal pour la période, seulement un quart environ des cahiers survivants ont cette forme ; les autres cahiers sont constitués de feuillets dépareillés, de bifeuillets ou d'un nombre variable de feuillets allant jusqu'à vingt. L'ordre original des feuillets dans le manuscrit est également loin d'être clair, et il est presque certain que l'ordre actuel des feuillets n'est plus du tout celui qu'il était à l'origine, un point souligné par au moins quatre foliotations différentes qui sont encore clairement visibles dans le manuscrit.

Comme cela a déjà été suggéré, il semble probable que cette complexité résulte au moins en partie de la manière dont l'enquête a été menée. Ainsi, Neil Ker a suggéré que la structure inhabituelle et les nombreux changements de main indiquaient que le parchemin était à l'origine préparé et réglé de façon cohérente, mais que les entrées étaient écrites « par n'importe quel scribe qui se trouvait être disponible » et que « la règle de base... semble avoir été de fournir au moins un cahier pour chaque tenant en chef, la taille minimale du

---

10. Pour l'infrastructure Archetype voir <https://archetype.ink>.



cahier étant probablement un seul bifeuillet »<sup>11</sup>. Il a également suggéré que les scribes avaient de vagues instructions, telles que « prenez un bifeuillet ; lorsque vous avez rempli le premier feuillet, ne passez pas au deuxième feuillet, mais prenez un autre bifeuillet et continuez sur cela, et ainsi de suite, jusqu'à quatre bifeuillets »<sup>12</sup>. Une telle hypothèse a des conséquences immédiates pour les questions concernant l'enquête elle-même : par exemple, elle suggère fortement que les scribes travaillaient à partir de matériaux organisés différemment de ceux de l'EDB, et probablement qu'ils travaillaient rapidement, avec des ressources importantes, et que la rapidité était jugée plus importante que la présentation. Cependant, l'un des objectifs clés du projet était de vérifier et (idéalement) de développer l'hypothèse de Ker ou d'en proposer une autre.

Compte tenu de la complexité du matériau, le besoin s'est rapidement fait sentir de trouver un moyen de gérer les informations codicologiques afin que l'équipe travaillant sur le projet (et tout le monde) puisse tester différentes hypothèses concernant la compilation du livre. Le plus sûr moyen d'y parvenir était d'utiliser l'outil informatique, mais une question beaucoup plus difficile était de savoir exactement comment les ordinateurs devaient être utilisés. Au cours des dix à quinze dernières années, le domaine de la recherche désormais appelé « paléographie numérique » s'est développé très rapidement<sup>13</sup>. En revanche, l'utilisation des outils numériques dans la modélisation et l'analyse des données codicologiques a été relativement moins étudiée. Beaucoup de travail a également été fait sur l'analyse automatique de la mise en page, mais moins sur le traitement automatique ou numérique de la structure des cahiers, la reliure, et autres aspects de la codicologie<sup>14</sup>. Cette différence pourrait s'expliquer en partie par le fait que, pour l'analyse automatique de la paléographie et de la mise en page, la seule entrée nécessaire pour le logiciel soit des images numériques de pages, qui sont maintenant disponibles en abondance. En revanche, pour d'autres questions de la codicologie, telles que la structure, l'information ne peut pas être déduite des images et se fonde donc plutôt sur des données saisies à la main ou récoltées à partir de catalogues très hétérogènes. Ce type de travail n'est pas nouveau : la codicologie quantitative est un domaine actif depuis plusieurs décennies, et les outils numériques ont certainement été utilisés pour cela, SfarData étant l'un des exemples les plus connus<sup>15</sup>. En termes de modélisation numérique et d'analyse de la structure des livres, des exemples importants sont ceux de Ligatus, une

11. « By any scribe that happened to be available » ; « the guiding rule... seems to have been to provide at least one quire for each holder, the minimum quire size being probably a single bifolium » : KER, 2002, vol. 2, p. 806.

12. « Take a bifolium ; when you have filled the first leaf do not go on to the second leaf, but take another bifolium and continue on that, and so on, up to four bifolia » : KER, 2002, vol. 2, p. 806.

13. STOKES, 2015a.

14. Le traitement automatique de la mise en page est le sujet d'une recherche intensive dans le domaine de *Historical Document Analysis*. Par exemple, le colloque *International Conference in Document Analysis and Recognition (ICDAR)* de 2019 aura une session entière dédiée à l'analyse automatique de la mise en page : <https://icdar2019.org/program/>.

15. SfarData ; BEIT-ARIÉ 1994. Pour la codicologie quantitative voir en particulier ORNATO 1997, ainsi que BOZZOLO et ORNATO, 1980.

unité de recherche de l'University of the Arts London, dont le projet inclut le « Language of Bindings Thesaurus », un outil de référence pour la terminologie de la reliure ; une thèse de doctorat sur la visualisation des structures de la reliure ; et au moins deux schémas XML pour décrire la disposition des feuillets et des cahiers<sup>16</sup>. L'outil le plus important à ce jour est peut-être VisColl et l'ensemble d'outils développé principalement par Dot Porter et Alberto Campagnolo, qui est décrit plus loin<sup>17</sup>.

Tous les cas énumérés ici se concentrent sur la modélisation et la présentation de la structure du manuscrit, plutôt que sur des approches plus automatisées. Cette méthodologie plus visuelle (même si aussi computationnelle) a également été appliquée à l'EDB. En effet, un principe clé qui a été décidé très tôt était que la meilleure approche n'était pas de tenter de donner automatiquement des réponses automatiques finales, mais plutôt d'utiliser le logiciel pour permettre l'exploration, l'analyse et la communication de preuves et de résultats pertinents<sup>18</sup>. En d'autres termes, il aurait certainement pu être possible de développer un système entièrement automatique pour tester différentes séquences de feuillets sans intervention humaine, en les comparant au moyen de critères tels que la continuité de réglure, la continuité du texte, etc., pour trouver la suite la plus probable. En effet, quelques expériences ont été menées autour de cette possibilité. Cependant, un algorithme comme celui-ci nécessite la mise en œuvre d'un grand nombre de paramètres complexes : par exemple, les experts du domaine devaient indiquer à l'ordinateur quelle importance relative donner aux différentes caractéristiques, comme décider que l'ordre des feuillets aurait un coefficient de 0,8 par rapport à la continuité de la réglure qui en aurait un de 0,6, et que le texte aurait peut-être une importance évaluée à un coefficient de 1,5, et ainsi de suite. En pratique, il était difficile pour les codicologues ou les historiens de choisir ces chiffres, ou même d'évaluer les différents résultats produits par la mise en œuvre de l'algorithme et de décider lesquels étaient les plus probables. On a donc plutôt utilisé l'ordinateur autant pour visualiser et communiquer les données codicologiques complexes de manière à les rendre compréhensibles et interrogeables par les membres de l'équipe et d'autres chercheurs : en d'autres termes, l'accent a été mis sur la visualisation numérique, l'interactivité et l'interface.

En pratique, les approches qui ont été développées pour le projet étaient fortement basées sur une modélisation préexistante de la production et de la circulation des « unités » codicologiques, développée par Andrist, Canart et Maniaci<sup>19</sup>. Ils commencent en définissant les « principales discontinuités observables dans le codex », telles que discontinuités de main, de texte, de parchemin,

---

16. PICKWOAD et al. [n.d.] ; CAMPAGNOLO, 2015 ; BRÜNING, HENZEL et PRAVIDA, 2013 ; PORTER et CAMPAGNOLO, 2014-2018

17. PORTER et CAMPAGNOLO, 2014-2018.

18. Pour des discussions de ce principe voir, par exemple, CLEMENT *et al.*, 2008 ; SCULLEY et PASANEK, 2008 ; et, pour la valeur de la visualisation interactive dans la création de connaissances, JESSOP, 2008a et notamment JESSOP, 2008b.

19. ANDRIST, CANART and MANIACI, 2013, en particulier chap. 2 et 3.

et ainsi de suite. Ils continuent en expliquant que les discontinuités sont d'abord envisagées de manière individuelle :

Nous relevons d'abord, de façon acritique, toutes celles d'un même type ; puis tentons de les regrouper en vue de faire ressortir celles d'entre elles qui sont significatives. Pour désigner clairement ces deux niveaux, nous appelons « éléments » les parties délimitées par les discontinuités les plus simples et « unités » les regroupements d'éléments.<sup>20</sup>

Bien qu'il soit décrit uniquement en version imprimée et pas du tout en termes numériques, le modèle est présenté très clairement et se traduit facilement en logiciel. En bref, il présente une taxonomie des unités de production (UniProd) et des unités de circulation (UniCirc), ainsi que des modèles pour la transformation des codices (en ajoutant ou en supprimant des matériaux), ou encore pour les discontinuités codicologiques. Par exemple, une unité de production de support matériel pourrait signifier que les cahiers ont été produits en un seul processus, de sorte que le parchemin, le cadre d'écriture, les lignes par page, etc. sont tous cohérents et ont probablement été produits de la même façon. Cependant, cela peut ou non correspondre à une unité de contenu, à savoir que le contenu écrit forme un ensemble cohérent. Pareillement, même si un livre peut comprendre une seule unité de production à la fois en matériel et en contenu, il ne s'ensuit pas nécessairement que ce livre a circulé dans son ensemble et n'a pas été divisé en parties et/ou a reçu des ajouts ou d'autres textes. Le modèle s'étend aussi aux « principales discontinuités observables dans un codex ». Par exemple, dans le cas de l'EDB, nous avons des « matières mélangées au hasard », les cahiers comportent une « série à changements aléatoires », mais il existe des « unités de marques de succession » de type « signes progressifs (numéros de cahiers, de feuillets, etc.) » et ainsi de suite.<sup>21</sup> Enfin, et surtout, ils fournissent un moyen de visualiser ces continuités sous des formes tabulaires qui contribuent de manière significative à l'identification des unités de production et de circulation.

## Analyse numérique

Inspirés par ces tableaux et leur efficacité dans la gestion de données codicologiques complexes, nous avons entrepris de produire des visualisations comparables pour l'EDB. L'équipe avait déjà décidé d'utiliser le logiciel « Archetype » pour le site web et les images, le texte et sa traduction, et les analyses paléographiques. « Archetype » est un logiciel libre, gratuit et « Open Source » développé à King's College de Londres. Il a été conçu à l'origine pour le projet « DigiPal », dans le but d'« offrir une méthodologie visible et reproductible permettant d'explorer les données paléographiques et de communiquer preuves et arguments à un public

---

20. ANDRIST, CANART and MANIACI, 2013, p. 83.

21. Ces expressions viennent de ANDRIST, CANART and MANIACI, 2013.

plus large »<sup>22</sup>. À la suite de « DigiPal » et grâce aux projets « Models of Authority » et « Exon Domesday », l'équipe de King's College de Londres a ajouté au logiciel la capacité de gérer et afficher des textes électroniques (telles qu'éditions et traductions), et d'associer un morceau de texte à une zone d'image<sup>23</sup>. « Archetype » fournissait donc déjà la plupart des fonctionnalités nécessaires pour « Exon Domesday », et il s'agissait donc d'ajouter une personnalisation supplémentaire pour la modélisation codicologique.

Le noyau de notre analyse numérique de la structure de l'EDB est la « Palaeographical and Codicological Description » du manuscrit qui a été écrite par Francisco Álvarez López<sup>24</sup>. Il a fallu examiner et décrire chaque page du manuscrit en quelques mois. Un nouveau formulaire d'encodage a été ajouté à l'éditeur de texte d'Archetype pour faciliter la saisie en ligne de ces descriptions. Ce modèle comprend du texte libre mais avec un balisage XML pour marquer les détails clés, ainsi le nombre de lignes par page, les dimensions de la page, etc. Ce système garantit une structure et une cohérence de base aux descriptions sans pour autant sacrifier la flexibilité éditoriale que requiert un manuscrit d'une telle irrégularité. Une fois la description terminée, le balisage permet d'extraire des détails paléographiques et codicologiques qui peuvent ensuite être ajoutés à la base de données sous une forme structurée pour analyse ultérieure. En pratique, plusieurs visualisations différentes ont été développées pour présenter ces données, chacune d'entre elles mettant l'accent sur différentes unités codicologiques. La raison en est double : d'une part, une seule présentation de toutes les unités présente trop d'informations à gérer pour les chercheurs, d'autre part, cette approche plus « pluraliste » nous permet de tester différentes représentations et différentes formes d'interactions pour voir lesquelles pourraient être les plus utiles.

### *Unités d'écriture et unités des mains*

La première visualisation développée avec ces données était plus paléographique que codicologique, car elle a été conçue pour représenter la distribution de l'activité des scribes dans tout le manuscrit. L'on sait depuis longtemps qu'environ vingt ou vingt-cinq scribes différents ont contribué à écrire l'EDB et que ces scribes interagissaient de manière complexe, certains semblant contribuer davantage à écrire le texte de base et d'autres se consacrer plus à des corrections et

22. « To offer a visible and replicable methodology to allow people to explore palaeographical data and to communicate their evidence and accompanying argument to a wider audience » : BROOKES *et al.*, 2015, p. 26. Le projet « DigiPal » a été financé au titre du septième programme-cadre de l'Union européenne (7<sup>e</sup> PCRD) sous la convention de subvention n° 263751 : <https://www.digipal.eu>. Pour l'infrastructure Archetype voir <https://archetype.ink> et <https://github.com/kcl-ddh/digipal>.

23. STOKES *et al.*, 2016; STOKES *et al.* 2019. « Models of Authority » est le résultat d'un financement du Arts and Humanities Research Council (AHRC) no. AH/L008041/1 : <https://www.modelsofauthority.ac.uk>.

24. ÁLVAREZ LÓPEZ, 2018.

des annotations<sup>25</sup>. Les questions à poser ici incluent donc de savoir quels scribes, le cas échéant, étaient plus concentrés sur l'écriture du corps du texte et lesquels intervenaient davantage sur les annotations ou les corrections; s'il était possible d'identifier des groupes de scribes qui semblaient souvent travailler ensemble; si certains scribes ou groupes de scribes étaient plus actifs dans certaines sections du manuscrit, et ainsi de suite. Afin d'aider à répondre à ces questions, ainsi que de contrôler plus facilement la qualité et la cohérence des données, nous avons développé une visualisation qui présente le manuscrit sous la forme d'une grille, chaque ligne verticale correspondant à une page et chaque ligne horizontale correspondant à un scribe. Pour les pages sur lesquelles un scribe donné était actif, une ligne verticale noire est affichée au point correspondant de la grille. Passer la souris sur une page ouvre une fenêtre contextuelle avec un résumé de l'activité des scribes sur cette page : ceci est important parce que les lignes noires ne font pas la différence entre les types d'activité (écrire le texte principal, les annotations, ou corrections). Au sommet de la visualisation en vert se trouve une représentation du nombre de scribes qui ont écrit sur cette page : plus les scribes sont nombreux, plus la ligne verticale est haute. Certaines lignes ont également été ajoutées en bas pour aider l'équipe du projet à voir et vérifier rapidement des détails tels que les pages décrites à ce jour, les images de haute qualité, etc. Le résultat est montré dans la figure 1, où on peut voir immédiatement, par exemple, que certains scribes – tels que ceux connus comme Upsilon et Zeta – avaient tendance à écrire des blocs de texte courts mais résultant d'un travail intense sans l'aide d'autres scribes; qu'Omicron a écrit des contributions relativement courtes qui se trouvent presque toutes sur les fol. 29-61v; qu'Alpha et Bêta étaient actifs dans une grande partie du manuscrit; et ainsi de suite<sup>26</sup>. Parfois, la signification de ces phénomènes devient plus claire en comparant avec le contenu et la structure du manuscrit. Par exemple, la contribution d'Omicron coïncide presque exactement avec la section du manuscrit consacrée au Dorset; il commence aussi sur la première page du cahier 6 et s'étend jusqu'à la dernière page du cahier 12. Les seules parties concernant le Dorset qui s'étendent avant et après la contribution d'Omicron sont, d'une part, celles qui concernent la description du *Demesne* du roi, écrite entièrement par Bêta sur quatre feuillets désormais dépareillés qui composent le cahier 5; d'autre part, ce sont celles qui concernent les terres de Gautier de Claville : celles-ci occupent un feuillet dépareillé qui suit immédiatement le cahier 12 et ont été écrites par Êta, un scribe qui a également contribué aux pages écrites par Omicron. La valeur aberrante dans la contribution d'Omicron, qui est clairement visible dans le tableau, est la mention de la seule exploitation de Gaucèlme d'Exeter sur ce qui est maintenant le fol. 398; le reste de cette page est vide en dehors d'une courte

25. FINN, 1964, p. 30-32; KER, 2002, vol. 2, p. 806 et 807; FLIGHT, 2006, ch. 5.

26. Nous suivons FLIGHT, 2006, ch. 5, en faisant référence aux scribes par les noms des lettres grecques transcrites en latin, avec l'ajout d'une main supplémentaire non identifiée par Colin Flight. Au moment de la rédaction de ce texte, nous nous référons à cette main supplémentaire comme U5 parce que c'était la cinquième d'une séquence de mains non identifiées à l'origine.

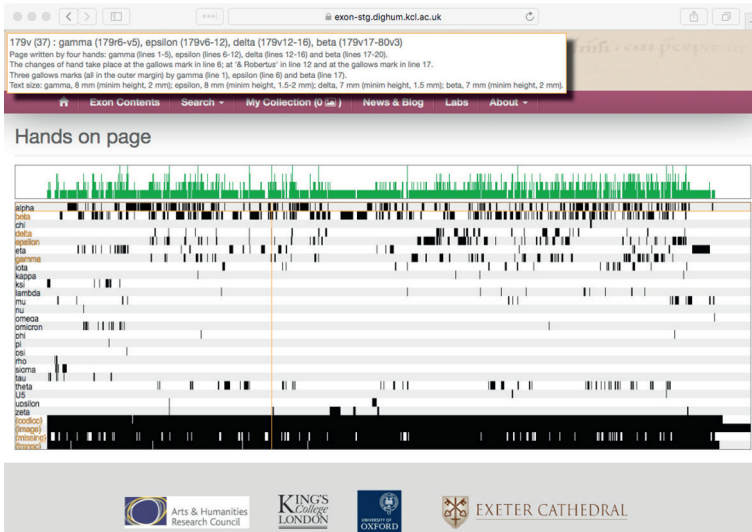


Fig. 1 : Visualisation de l'activité des scribes à travers l'EDB.

L'encadré en haut de la capture d'écran montre un résumé de l'activité des scribes pour la page sélectionnée, fol. 179, dont la position dans le manuscrit est indiquée par la ligne jaune verticale. La table est disponible à

<http://www.exondomesday.ac.uk/labs/distribution-of-scribal-hands/>

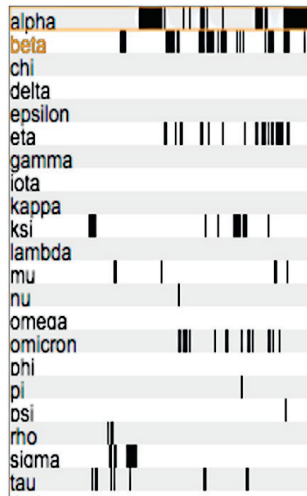


Fig. 2 : Détail de la visualisation représentée sur la figure 1.

La boîte contient les fol. 25-62v, à savoir les cahiers 5-13, qui correspondent aux terres du Dorset.

note que *Eduuardus tenet .iii. hidas terrae*, et le verso est également vide. Il s'agit d'un feuillet dépareillé, situé actuellement à la fin du cahier 77. Bien qu'il s'agisse d'un petit exemple, il illustre plusieurs points théoriques qui ont été soulevés

plus haut : l'activité de ce scribe est contenue dans des unités de production codicologiques, comme cela est indiqué par les coïncidences dans les ruptures de l'activité des scribes et des cahiers, et cela peut également suggérer qu'il y a plusieurs unités distinctes de production : le domaine du roi dans le Dorset dans le cahier 5 (fol. 25-28v), le reste du Dorset dans les cahiers 6-13 (fol. 62v), et dans le cahier 77 (fol. 398r-v).

### *Unités de réglure, unités de mise en page*

Comme le montre ce qui précède, la visualisation de la distribution des mains des scribes permet de voir des unités de production « UniMain » selon le schéma développé par Andrist, Canart et Maniaci. Cependant, comme cela le montre aussi, les mains des scribes ne suffisent pas en elles-mêmes, car c'est la coïncidence de différentes unités qui nous indique où le processus de production était susceptible d'avoir été différent. À cette fin, il est également nécessaire de visualiser d'autres unités telles que les limites des cahiers, les changements de taille de page, les changements de piqûre et de réglure, etc., et cela est le sujet de la deuxième visualisation que nous avons développée pour le projet Exon. Le résultat est un tableau similaire à celui proposé par Andrist, Canart et Maniaci, mais adapté au projet. Chaque ligne du tableau représente une page du manuscrit et les lignes horizontales pleines représentent les limites du cahier. Les différentes colonnes montrent des données extraites automatiquement de la description paléographique et codicologique balisée. La colonne la plus à droite donne une miniature de la page correspondante que l'on peut agrandir avec un passage de la souris. Cette colonne offre donc un aperçu photographique de l'intégralité du manuscrit, page par page.

L'un des défis ici est de déterminer les limites des unités codicologiques : par exemple, étant donné que la réglure est quelque peu irrégulière, à quel point la taille d'une trame doit-elle être différente pour constituer une nouvelle unité ? De plus, qu'est-ce qui constitue une nouvelle unité textuelle dans un manuscrit comme EDB ? Une entrée différente ? Un tenant en chef différent ? Un comté différent ? Les réponses à ces questions sont nécessaires pour créer un tableau du genre spécifié par Andrist, Canart et Maniaci, mais en même temps ce sont précisément les questions auxquelles l'équipe de recherche tente de répondre. Pour cette raison, entre autres, la table a été légèrement adaptée, et un détail du résultat est illustré par la figure 3.

La coloration rouge est générée automatiquement et vise à indiquer les discontinuités dans les catégories concernées : plus la couleur est foncée, plus grande est la discontinuité. Pour l'exemple montré par la figure 3, dans les colonnes « largeur de page » et « largeur de réglure » (*Page Width* et *Written Width*), on voit des variations relativement limitées surlignées en rose. En revanche, les feuillets 134r et 135r montrent tous les deux une variation significative par rapport aux feuillets qui les précèdent dans la hauteur du cadre de réglure, le nombre de lignes réglées par page, et la forme de la piqûre (tels que « y-shaped », « star-shaped », et ainsi de suite). L'examen montre que 135r revient

Link	Order	Quire	Locus (Barnes)	Locus (top right)	Conjoint	Page Height	Page Width	Written Height	Written Width	Rul. Height	Rul. Width	Rul. Count	Parch. Side	Perf. Shape	Thumbr
<a href="#">View</a>	264	28	132v	None	125	264	171	193	117	187	114	20	Hair	y-shaped	
<a href="#">View</a>	265	29	133r	135	138	261	176	192	135	185 (0.01)	134	20	Hair	y-shaped and star-shaped	
<a href="#">View</a>	266	29	133v	None	138	261	176	192	135	185	134	20	Flesh	y-shaped and star-shaped	
<a href="#">View</a>	267	29	134r	146	137	263	164	176	128	113	121	12	Flesh	y-shaped	
<a href="#">View</a>	268	29	134v	None	137	263	164	203	128	113	121	12	Hair	y-shaped	
<a href="#">View</a>	269	29	135r	137	136	263	172	193	143	186	133	20	Flesh	y-shaped and star-shaped	
<a href="#">View</a>	270	29	135v	None	136	263	172	183	137	186	133	20	Hair	y-shaped and star-shaped	
<a href="#">View</a>	271	29	136r	138	135	263	173	73	142	188	134	20	Hair	y-shaped and star-	

Fig. 3 : Un extrait de la «vue codicologique» du projet Exon Domesday, montrant la dernière page du cahier 28 (132v) et la première moitié du cahier 29. Le rouge indique une discontinuité dans la mesure respective. La table complète est disponible à <http://www.exonduomesday.ac.uk/labs/codicological-visualisation/>

en fait à la même réglure que celle de 133rv, ce qui indique que la piqûre et la réglure du feuillet 134 sont significativement différentes de celle des feuillets situés avant et après, et ont donc vraisemblablement été piqués et réglés par une personne différente, ou à un moment ou à un endroit différents. Ce bifeuillet aurait très bien pu être ajouté à ce cahier après l'assemblage des feuillets initiaux, comme semble aussi le confirmer la rupture de la «règle de Gregory» du côté chair face chair entre 134v et 135r, discontinuité également mise en évidence dans la visualisation présentée ici<sup>27</sup>. Cette manière de faire ressortir les points où coïncident plusieurs formes de discontinuités facilite considérablement l'identification des ruptures dans la production et la circulation, ce qui a permis de répondre aux questions posées ci-dessus.

### Unités de cahiers

Comme on l'a déjà dit, chaque ligne de la visualisation codicologique décrite ci-dessus correspond à une page. En effet, l'unité clé du livre en numérisation est la page : c'est-à-dire, pour des raisons simplement liées au flux de travail

27. La «règle de Gregory» est une «règle selon laquelle, dans un cahier de parchemin, les pages qui se font face présentent la même face du parchemin, c'est-à-dire le côté chair ou le côté poil, alternativement». MUZERELLE, 2003, III.1, 313.15a.



et à son efficacité, les livres sont normalement numérisés en photographiant d'abord toutes les pages recto et ensuite tous les versos. Cette centralité de la page est évidente dans la plupart des outils de visualisation de manuscrits ou de livres, tels que ceux qui sont utilisés par la British Library, la Bodleian d'Oxford, le visualiseur « Mirador », et ainsi de suite. Cependant, en lisant un livre, l'unité principale n'est pas la page mais plutôt l'ouverture en double-page, en visualisant les deux pages opposées. Ceci est de plus en plus reconnu, et en effet les visualiseurs permettent maintenant souvent de voir les deux pages en regard (par ex. le « Book View » des versions récentes de Mirador), bien que l'efficacité de ces visualisations varie considérablement. Cependant, en termes de structure physique du livre, l'unité primaire n'est ni la page ni l'ouverture mais plutôt le bifeuillet, à savoir la feuille de parchemin qui est pliée en deux pour constituer deux contre-feuillets dans le livre. Cette idée est la clé de VisColl, une boîte à outils qui a été développée par Dot Porter, Alberto Campagnolo et d'autres<sup>28</sup>. Il comprend un schéma XML pour décrire les cahiers et feuillets, ainsi qu'un « Codicological Modeller » qui comprend un formulaire en ligne dans lequel on entre la structure du livre et on reçoit une description de la structure encodée en XML comme résultat. L'élément principal dans cette suite logicielle est VisColl lui-même, qui prend la description XML et une liste de références aux images numériques de chaque page dans le manuscrit, et génère ensuite plusieurs sorties dont une formule de classement générée automatiquement et un ensemble complet de diagrammes de cahiers pour l'ensemble du manuscrit. Il comprend également une « déconstruction virtuelle » du volume, montrant le recto et le verso de chaque bifeuillet tel qu'il apparaîtrait si le manuscrit était dissocié. Cela permet aux chercheurs de voir les feuilles originales dépliées, par opposition aux doubles pages du manuscrit. Il est particulièrement utile pour comprendre la construction du livre, car il permet de repérer des feuilles qui ont été préparées différemment de celles qui les entourent, que ce soit en qualité de parchemin, cadre de réglure, lignes par page, etc.

Malheureusement, la quantité des pages et la complexité de la structure de l'EDB le rendent inutilisable pour le « Codicological Modeller », car il faudrait beaucoup de temps pour entrer les données dans les formulaires en ligne. Cependant, il est possible de partir de la visualisation codicologique décrite ci-dessus et d'écrire des scripts XSLT pour la transformer en XML sous la forme requise par VisColl ; un extrait du résultat est montré dans la figure 4<sup>29</sup>. Cela fait immédiatement ressortir plusieurs points. Tout d'abord, cela a permis de clarifier certaines fautes mineures dans la description codicologique et a ainsi aidé à vérifier et à corriger ce document extrêmement long et complexe.

---

28. PORTER et CAMPAGNOLO, 2014-2018.

29. En pratique, l'automatisation complète n'est pas simple en raison de la façon dont les feuillets dépareillés sont gérés sous XML ; donc au moment de l'écriture, le XSLT génère une sortie très proche de ce qui est requis et les corrections finales sont faites à la main. Ce processus semi-automatisé est suffisant pour les besoins de ce projet, mais pourrait être entièrement automatisé et pourrait même être intégré dans l'architecture Archetype si d'autres projets l'exigeaient.



Fig. 4 : Cahier 79 de l'EDB sous forme de « déconstruction virtuelle » créé avec VisColl. La « déconstruction » complète est disponible à <http://exonduomesday.ac.uk/lab/viscoll/>).

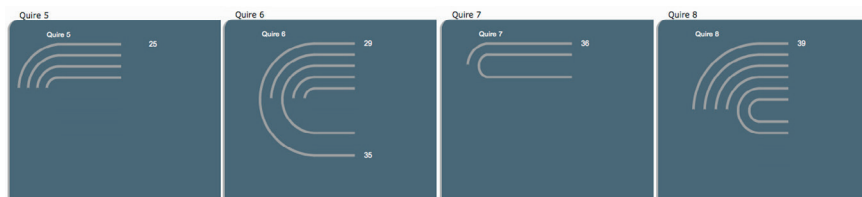


Fig. 5 : Extrait de diagrammes de cahiers générés avec VisColl.

Les diagrammes de cahiers qui en résultent pourraient également être vérifiés, parce que le manuscrit a été désassemblé avant une éventuelle restauration et nouvelle reliure, et au cours de ce processus le conservateur a préparé ses propres diagrammes à la main. Une comparaison a donc été faite entre les deux séries de diagrammes pour vérifier l'une par rapport à l'autre. Plus important encore, cependant, cette façon de représenter la structure physique du livre manuscrit s'est déjà révélée très utile pour comprendre l'objet et la façon dont il a été composé. Pour ceux qui n'ont pas la possibilité de passer beaucoup de temps à travailler étroitement avec l'original, des visualisations de ce type fournissent également des informations importantes.

### *Unités de contenu*

Le dernier type d'unité à aborder ici est celui du contenu textuel. On a souhaité pouvoir accéder au contenu par deux outils différents dans le projet Exon, afin d'avoir deux points de vue différents. Le premier est une table des matières simple qui liste les unités textuelles principales selon l'ordre dans lequel elles apparaissent actuellement dans le manuscrit. Bien que cela puisse sembler

banal, et cela l'est effectivement du point de vue informatique, son importance ne doit pas être négligée car cette visualisation fournit un aperçu très commode et toujours extrêmement utile des principales unités de contenu textuel. Ces unités textuelles se rapportent aux terres qui sont décrites : dans la plupart des cas, elles correspondent simplement au tenant en chef respectif. En pratique, cela donne un tableau simple en quatre colonnes principales comprenant les numéros d'entrées, le contenu de la section, le type d'entrée (comptes de *geld*, *boroughs*, *fiefs*, *terrae occupatae* et ainsi de suite) et le comté<sup>30</sup>. Le deuxième outil développé pour comprendre les unités de contenu textuel est beaucoup plus expérimental à ce jour. Comme il a été mentionné ci-dessus, l'EDB a été relié plusieurs fois tout au long de sa vie, et la séquence des cahiers a changé de manière significative au cours de ces nouvelles reliures. Une question importante est donc de savoir quel ordre les cahiers avaient dans le manuscrit quand il a été produit pour la première fois, si du reste il était même relié en un seul volume plutôt que gardé comme des ensembles de cahiers. Pour aider à cela, il serait utile de disposer d'un outil qui permette aux chercheurs de changer dynamiquement l'ordre des cahiers ou même des feuillets et de faire en sorte que les différents diagrammes et même le texte lui-même soient mis à jour. On peut encore approfondir cette idée, car toutes les séquences de cahiers ou de feuillets ne sont pas physiquement possibles, et certaines sont plus probables que d'autres. Certaines contraintes peuvent nous paraître triviales, mais il est nécessaire de les préciser pour l'ordinateur, comme le fait que chaque feuillet comporte deux côtés, un recto et un verso. D'autres, cependant, sont également évidentes mais peut-être plus complexes, comme le fait que deux feuillets conjoints qui forment actuellement un bifeuille (sans aucun assemblage) doivent toujours avoir formé ce même bifeuille; de même, dans un tel bifeuille, si le recto du premier feuillet est côté poil, le verso du feuillet conjoint doit aussi être côté poil, et ainsi de suite. De cette manière, on peut construire une liste de contraintes « dures » qui spécifient les dispositions qui ne sont pas physiquement possibles et doivent donc être exclues de toute réorganisation du document. De même, d'autres configurations de pages sont physiquement possibles mais peu probables en pratique. Par exemple, il est très rare qu'un bifeuille soit retiré, plié « en arrière » puis réinséré dans le manuscrit, de sorte que ce qui était le recto du premier feuillet devienne le verso du deuxième feuillet, et ainsi de suite<sup>31</sup>.

Des règles comme celles-ci peuvent être rédigées relativement facilement, même si une réflexion méthodique et prudente est nécessaire<sup>32</sup>. La question est alors de savoir comment elles peuvent être mises en pratique. Une approche

30. *Geld*, *hide* : voir n. 5 ci-dessus. *Boroughs* : des bourgs. *Terrae occupatae* : « les terres sous occupation ». Ceci est un titre « fourre-tout » qui couvre plusieurs sortes d'illégalité, par exemple la saisie des terrains d'autrui et l'aliénation de certains obligations, droits et privilèges.

31. Bien que très rare, ce phénomène a été parfois observé : deux exemples à notre connaissance sont Cambridge, Corpus Christi Library, ms 367, et Oxford, Bodleian Library, ms auct. f.4.32. Pour la discussion concernant le premier manuscrit voir TREHARNE, 1998, 2000 et 2010 ; pour le dernier voir HUNT, 1961.

32. Pour un premier essai d'établir une telle liste voir STOKES, 2015b.

possible consiste à utiliser une combinaison du modèle documentaire fourni par la « Text Encoding Initiative » avec un schéma personnalisé. La TEI est bien connue comme la norme *de facto* pour les éditions numériques. Actif depuis plus de trente ans maintenant, le consortium a longtemps fourni des directives pour le balisage de textes de nombreux types et formes. Jusqu'à une époque relativement récente, il s'est concentré sur le texte en tant que séquence de mots et a accordé relativement peu d'attention à sa forme matérielle, à savoir la représentation du document, sa mise en page, etc. Cela a changé quand une nouvelle option a été ajoutée pour fournir des éditions « document-centric » qui se concentrent sur le document physique et sa mise en page<sup>33</sup>. Plus précisément, on peut spécifier la surface sur laquelle un texte apparaît, et on peut grouper les surfaces de façon hiérarchique. Un tel modèle est facilement applicable à un contexte tel que le nôtre, où chaque face d'un feuillet peut être modélisée comme une surface; un feuillet est donc un groupe de deux surfaces, un cahier comprend un groupe de feuillets, et ainsi de suite. Le principe est illustré sur la figure 6 qui représente un cahier composé d'un bifeuillet renfermant deux feuillets dépareillés. L'élément <surfaceGrp> (groupe de surfaces) externe représente le cahier, à l'intérieur duquel se trouve un second <surfaceGrp> représentant le bifeuillet. Le bifeuillet lui-même comprend deux autres groupes de surfaces pour les deux feuillets, ainsi que les deux feuillets dépareillés qui à leur tour comprennent chacun un feuillet. Chaque feuillet contient deux éléments de <surface> représentant des pages, et chaque page contient alors le texte; ceux-ci ne sont pas représentés sur la figure pour des raisons de clarté et de concision mais sont indiqués par les crochets comme « [20 lignes] ». Chaque groupe de surfaces contient également un attribut @corresp qui pointe vers une section différente du document XML au moyen d'un identifiant unique qui est justement défini dans cette section du document. Cette section comprend des balises spécialisées développées pour le projet mais modélisées de façon proche de celles du modèle VisColl et illustrées dans la figure 7. Ici, chaque élément <exon:folio> correspond à un feuillet du manuscrit, les attributs donnant des informations sur la composition du feuillet : par exemple si le recto est côté poil et le verso est côté chair ou vice-versa (@hair="hf" ou @hair="fh" respectivement), avec quelle page (le cas échéant) il est conjoint, que ce soit au recto ou au verso, et ainsi de suite. Enfin, les contraintes codicologiques ont été encodées à l'aide de Schematron, qui est un langage de programmation standard permettant de spécifier des règles sur ce qui est autorisé ou non dans un codage XML<sup>34</sup>.

Bien qu'il ne soit pas possible dans les limites de cette expérience d'encoder toutes les contraintes codicologiques (il n'est d'ailleurs pas évident que toutes les contraintes puissent être encodées dans les limites de Schematron), néanmoins cette méthode a permis d'établir un ensemble de règles déjà important. Il était alors simple de charger le XML et Schematron dans un éditeur XML (dans

33. PIERAZZO, 2017; BRÜNING, HENZEL et PRAVIDA, 2013.

34. Le fichier Schematron a été écrit par Peter Stokes et est disponible à <https://gist.github.com/pastokes/e80dcd50c86be4351288>.

```

<surfaceGrp type="quire" corresp="#q1">
  <surfaceGrp type="bifolium">
    <surfaceGrp type="folio" corresp="#1a1"> [20 lines]
    <surfaceGrp type="singleton">
      <surfaceGrp type="folio" corresp="#1b1"> [19 lines]
    </surfaceGrp>
    <surfaceGrp type="singleton">
      <surfaceGrp type="folio" corresp="#2a1"> [14 lines]
    </surfaceGrp>
    <surfaceGrp type="folio" corresp="#2b1"> [15 lines]
  </surfaceGrp>
</surfaceGrp>

```

Fig. 6 : Extrait de l'encodage TEI XML utilisé pour encoder le texte du livre en fonction de sa mise en page en termes de pages, feuillets, bifeuillets et cahiers. Le texte lui-même n'est pas donné pour des raisons de concision.

```

<exon:quire n="1" xml:id="q1">
  <exon:folio xml:id="1a1" n="1" conjoins="2b1" single="false" ruled="before" hair="hf"/>
  <exon:folio xml:id="1b1" n="2" single="true" ruled="unruled" hair="fh"/>
  <exon:folio xml:id="2a1" n="3" single="true" ruled="after" hair="hf"/>
  <exon:folio xml:id="2b1" n="4" conjoins="1a1" single="false" ruled="before" hair="fh"/>
</exon:quire>
<exon:quire n="2">
  <exon:folio xml:id="3a1" n="1" conjoins="4b1" single="false" ruled="before" hair="hf"/>
  <exon:folio xml:id="3b1" n="2" conjoins="4a1" single="false" ruled="unruled" hair="fh"/>
  <exon:folio xml:id="4a1" n="3" conjoins="3b1" single="false" ruled="after" hair="hf"/>
  <exon:folio xml:id="4a1*" n="3*" single="true" ruled="" hair="hf"/>
  <exon:folio xml:id="4b1" n="4" conjoins="3a1" single="false" ruled="before" hair="fh"/>
</exon:quire>

```

Fig. 7 : Extrait du XML utilisé pour encoder la composition physique des feuillets.

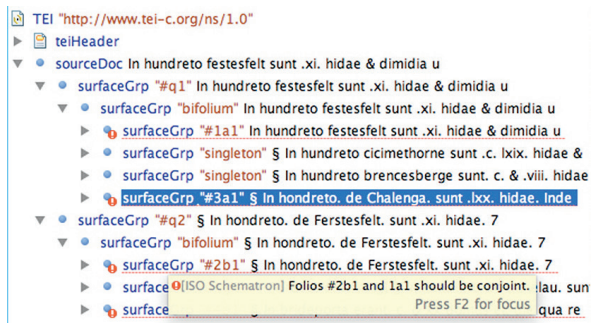


Fig. 8 : Exemple de réorganisation des pages à l'aide de l'éditeur XML Oxygen, avec la vue « Sommaire » dans le volet de gauche et le texte visible dans la partie droite. L'éditeur XML a identifié que l'arrangement actuel est en violation des contraintes et a donc alerté l'utilisateur en conséquence.

notre cas Oxygen). La visualisation « Sommaire » de l'éditeur a alors fourni une vue d'ensemble de la structure du manuscrit représentée en XML, avec laquelle il est possible de glisser et déposer des feuillets, bifeuillets et même des cahiers dans différentes positions les uns par rapport aux autres. Ce faisant, le texte de l'éditeur est automatiquement mis à jour pour refléter la nouvelle mise en ordre, mais si l'ordre viole l'une des contraintes, l'utilisateur est immédiatement averti. Un exemple de ceci est illustré à la figure 8. Cette utilisation de Schematron et Oxygen n'est pas habituelle et n'est pas forcément ce que leurs créateurs ont jamais imaginé, puisque le langage et l'outil sont conçus pour vérifier que la

structure d'un document XML est valide pour un schéma donné. Néanmoins, cette approche donne à l'équipe du projet un moyen très pratique de faire des expériences sans besoin de développer un logiciel personnalisé.

### Conclusions et développement futur éventuel

Le projet Exon Domesday a été l'occasion de développer plusieurs façons de visualiser à la fois le texte et la production de l'Exon Domesday Book, et de pouvoir interagir avec eux. Aucune vue n'est complète en elle-même et, en effet, aucune vue unique ne peut jamais fournir une compréhension complète d'un objet aussi complexe que celui-ci. Ce n'est cependant pas un problème : au contraire, l'un des buts de ces visualisations est justement de simplifier, de prendre en considération un élément trop complexe pour être compris en une seule fois et de se focaliser sur des aspects spécifiques de cet objet pour mieux l'analyser. C'est précisément le but de la modélisation que de créer un modèle qui soit une simplification de l'objet dans le but de le rendre malléable à fin d'analyse, d'éliminer le « bruit » et de permettre au chercheur de se concentrer sur un élément précis à un moment donné<sup>35</sup>. Il existe néanmoins des moyens évidents d'étendre et d'améliorer les visualisations. L'un d'eux serait de permettre la représentation de toutes les différentes combinaisons d'unité de production et de circulation, mais, comme l'ont clairement indiqué Andrist, Canart et Maniaci, la signification de ces unités réside dans leur combinaison et les discontinuités qui les séparent. Cela signifie qu'en pratique il faut évoluer entre toutes les visualisations présentées ci-dessus pour comprendre quelles sont les unités de production et donc déterminer quels sont les ordres de feuillets et de cahiers qui sont les plus probables. En principe, réaliser ceci ne serait pas difficile, dans la mesure où les informations déjà présentes sur le site sont suffisantes pour identifier les points où coïncident les frontières des textes, manuscrits, matériaux et cahiers, mais il est moins évident de savoir comment présenter toutes ces informations d'une manière facilement compréhensible. Une autre approche possible, avec laquelle nous avons fait quelques expériences à ce jour, consiste à prendre une approche plus computationnelle. Comme il a été dit, le but de la composante numérique de ce projet était de fournir des moyens interactifs de visualisation et de manipulation du matériau de sorte que les chercheurs puissent réaliser des expériences, tester des hypothèses, et de manière générale créer et communiquer des connaissances grâce à ces analyses et à ces interactions. Une approche complémentaire consiste à utiliser des outils plus automatisés et computationnels. Par exemple, il a déjà été observé que certaines mises en ordre de feuillets sont impossibles, d'autres sont possibles mais peu probables, alors que d'autres sont possibles et même au moins aussi probables que l'ordre actuel. En principe, alors, on pourrait rechercher avec l'outil informatique l'ensemble des mises en ordre possibles, en proposant que les mises en ordre que

---

35. MCCARTY, 2004, p. 255.

l'ordinateur calcule soient les plus probables sur la base des critères établis par l'équipe de recherche. L'un des défis à relever ici est de savoir comment décider de la probabilité des différents ordres, d'autant que l'EDB est très inhabituel dans sa construction et que certains arrangements jugés très improbables dans les manuscrits « normaux » étaient manifestement représentés dans l'EDB. Une expérimentation a donc été réalisée avec la « Vue Codicologique », dans laquelle on peut sélectionner des sections du manuscrit et modifier la pondération des différentes valeurs : par exemple, on peut décider d'accorder une priorité à la piquûre, mais considérer la continuité dans la taille de la page comme une faible priorité, et ainsi de suite. L'ordinateur calcule alors l'ordre des pages le plus probable en fonction de ces pondérations. Dans la pratique, cela n'a pas encore donné de résultats utiles, en partie parce qu'il est difficile de décider quelles pondérations donner à quelles valeurs, et en partie aussi parce que ce n'est bien sûr qu'une des manières possibles d'envisager le matériau avec un si grand nombre de permutations dans l'ordre des pages. Cependant, cela suggère une autre direction que pourraient prendre des recherches futures. Quoi qu'il en soit, ces outils ont déjà donné à l'équipe de nouvelles perspectives de recherche. En particulier, les historiens et les paléographes signalent que le site web leur est devenu indispensable dans leur travail et qu'ils l'utilisent quotidiennement pour consulter et comprendre les données. La recherche qui a résulté de l'utilisation de cet outil a déjà produit des résultats significatifs sur le manuscrit et sa production, résultats qui affinent et modifient notre compréhension de l'EDB et donc de l'enquête Domesday, ainsi que de la pratique documentaire de l'Angleterre anglo-normande précoce<sup>36</sup>.

## Bibliographie

### *Sources manuscrites*

*Exon Domesday Book*. Exeter, Exeter Cathedral, ms 3500 (vers 1086).

### *Sources numériques*

*Archetype*, London, King's College, 2017. <https://archetype.ink>.

*DigiPal: Digital Resource and Database for Palaeography, Manuscripts and Diplomatic*. Londres, King's College London, 2011–2014. <http://digipal.eu/>.

« Exon Domesday », Frank THORN (éd. et trad.), in *Exon Domesday: The Domesday Survey of South-West England*, Peter A. STOKES (éd.) (Studies in Exon Domesday I, Julia CRICK éd gen.), London, King's College, 2018. <http://www.exonduomesday.ac.uk>.

« The Friends of Exeter Cathedral », Exeter, Exeter Cathedral [n.d.]. <https://www.exeter-cathedral.org.uk/support-us/the-friends-of-exeter-cathedral/>.

36. BAXTER, 2017; ALVÁREZ LOPEZ et CRICK, à paraître; BAXTER *et al.*, en préparation.

- International Conference on Document Analysis and Recognition (ICDAR)*, Sydney, 2019. <https://icdar2019.org/program/>.
- Models of Authority: Scottish Charters and the Emergence of Government 1100–1250*. Londres, King's College, 2015–17. <http://www.modelsofauthority.ac.uk>.
- PICKWOAD, Nicholas, et al., *The Language of Bindings Thesaurus*, London, University of the Arts, [n.d.]. <http://www.ligatus.org.uk/lob/>.
- SfarData: The Codicological Data-Base of the Hebrew Palaeography Project*, The Israel Academy of Sciences and Humanities, [n.d.]. <http://sfardata.nli.org.il/>.

## Études

- ÁLVAREZ LÓPEZ, FRANCISCO, « Palaeographical and Codicological Description », in *Exon Domesday: The Domesday Survey of South-West England*, Peter A. STOKES (éd.) (Studies in Exon Domesday I, Julia CRICK éd. gen.), Londres, King's College, 2018. <http://www.exondomesday.ac.uk>.
- ALVÁREZ LOPEZ, FRANCISCO, et CRICK, Julia, « Decision-making and work flow in the making of Exon Domesday », in *Proceedings of the 20<sup>th</sup> Colloquium of the Comité international de paléographie latine*, Brepols, à paraître.
- ANDRIST, Patrick, CANART, Paul et MANIACI, Marilena, *La syntaxe du codex : essai de codicologie structurale*, Turnhout, Brepols, 2013 (Bibliologia 34).
- BAXTER, Stephen, « The Domesday controversy : A review and a new interpretation », *The Haskins Society Journal*, 29, 2017, p. 225-293. URL : [www.jstor.org/stable/j.ctv1ntjdp.14](http://www.jstor.org/stable/j.ctv1ntjdp.14).
- BAXTER, Stephen, LEWIS, Chris, CRICK, Julia, et THORN, Frank, *Making Domesday: The Conqueror's Survey in Context*, (Studies in Exon Domesday II, Julia CRICK éd. gen.), Oxford, Oxford University Press, en préparation.
- BEIT-ARIÉ, Malachi, « SfarData: The Henri Schiller Codicological Database of the Hebrew Palaeography Project, Jerusalem », *Gazette du livre medieval*, 25, 1994, p. 24-29. DOI : 10.3406/galim.1994.1280.
- BROOKES, Stewart, STOKES, Peter A., WATSON, Matilda et MARQUES DE MATOS, Debora, « The DigiPal Framework for European Scripts and Decorations », in *Writing Europe 500–1450: Texts and Contexts*, Aidan CONTI, Orietta DA ROLD, et Philip SHAW, (éd.), Cambridge, C.S. Brewer, 2015 (Essays and Studies 68), p. 25-58.
- BRÜNING, Gerrit, HENZEL, Katrin et PRAVIDA, Dietmar, « Multiple Encodings in Genetic Editions: the Case of 'Faust' », *Journal of the Text Encoding Initiative*, 4, 2013. DOI: 10.4000/jtei.697 <https://journals.openedition.org/jtei/697>.
- BOZZOLO, Carla et ORNATO, Ezio, *Pour une histoire du livre manuscrit au Moyen Âge. Trois essais de codicologie quantitative*, Paris, éditions du CNRS, 1980 (Textes et études, 2).
- CAMPAGNOLO, Alberto, *Transforming structured descriptions to visual representations. An automated visualization of historical bookbinding structures*, Thèse de doctorat, University of the Arts, Londres, 2015 (dactyl.).
- CLARKE, Howard B., « The Domesday Satellites », in *Domesday Studies : a re-assessment*, Peter SAWYER (ed.), Londres-Baltimore, Eward Arnold, 1985, p. 50-70.



- CLEMENT, Tanya, STEGER, Sarah, UNSWORTH, John et USZKALO, Kirsten, «How Not to Read a Million Books», 2008. URL : <http://www.people.virginia.edu/~jmu2m/hownot2read.html>.
- FINN, Rex W., *Domesday Studies: The Liber Exoniensis*, Londres, Longmans, 1964.
- FLIGHT, Colin, *The Survey of the Whole of England: Studies of the Documentation Resulting from the Survey Conducted in 1086*, BAR British Series 405, Oxford, Archaeopress, 2006.
- HUNT, Richard W., *St Dunstan's Classbook from Glastonbury: Codex Biblioth. Bodleiana Oxon. Auct. F.4.32*, Amsterdam, North-Holland Publishing Co., 1961 (Umbræ codicum occidentalium 4).
- JESSOP, Martyn, «The Inhibition of Geographical Information in Digital Humanities Scholarship», *Literary and Linguistic Computing*, 23/1, 2008, p. 39-50. DOI : 10.1093/llc/fqm041.
- JESSOP, Martyn, «Digital Visualization as a Scholarly Activity», *Literary and Linguistic Computing*, 23, 2008/3, p. 281-293. DOI : 10.1093/llc/fqn013.
- KER, Neil R., *Medieval Manuscripts in British Libraries*, Oxford, Clarendon Press, 2002, 5 vols.
- MCCARTY, Willard, «Modeling: A Study in Words and Meanings», in *A Companion to Digital Humanities*, Susan SCHREIBMAN, Ray SIEMENS et John UNSWORTH (éd.), Oxford, Blackwell, 2004, p. 254-270.
- MUZERELLE, Denis, *Vocabulaire codicologique : répertoire méthodique des termes français relatifs aux manuscrits, avec leurs équivalents en anglais, italien, espagnol, édition hypertextuelle*, Institut de Recherche et d'Histoire des Textes (IRHT), Paris, 2002-2003. Réédition dans *Codicologia*, IRHT, Paris 2011. URL : <http://codicologia.irht.cnrs.fr>.
- ORNATO, Ezio, *La Face cachée du livre médiéval : l'histoire du livre vue par Ezio Ornato, ses amis et ses collègues*. Rome, Viella, 1997.
- PIERAZZO, Elena, «Facsimile and Document-Centric Editing», in *Creating a Digital Scholarly Edition with the Text Encoding Initiative*, Marjorie BURGHART (éd.), 2017, chapitre 5. URL : <https://www.digitalmanuscripts.eu/digital-editing-of-medieval-texts-a-textbook/>.
- PORTER, Dot and CAMPAGNOLO, Alberto, *VisColl Collation Visualizer*, Philadelphie, University of Pennsylvania, 2014-2018. URL : <https://github.com/leoba/VisColl/>.
- SCULLEY, David, and M. PASANEK, Bradley, «Meaning and Mining: The Impact of Implicit Assumptions in Data Mining for the Humanities», *Literary and Linguistic Computing*, 23/4, 2008, p. 409-424. DOI : 10.1093/llc/fqn019.
- STOKES, Peter A., «Digital Approaches to Palaeography and Book History: Some Challenges, Present and Future», *Frontiers in Digital Humanities*, 2/5, 2015a. DOI : 10.3389/fdigh.2015.00005
- STOKES, Peter A., «Modelling Codicology I: Sequence in Gatherings, Folios and Pages», in *Exon Domesday: The Domesday Survey of South-West England*, Peter A. STOKES (éd.), Studies in Exon Domesday I, Julia CRICK (éd. gen.), London, King's College, 2015b. URL : <http://www.exondomesday.ac.uk/blog/modelling-codicology-i-sequence-in-gatherings-folios-and-pages/>

- STOKES, Peter A., BROOKES, Stewart, NOËL, Geoffroy, DAVIES, John R., WEBBER, Teresa, BROUN, Dauvit, TAYLOR, Alice et TUCKER, Johana, « The Models of Authority Project: Extending the DigiPal Framework for Script and Decoration », in *Digital Humanities 2016: Conference Abstracts*, Maciej EDER et Jan RYBICKI (éd.), Cracovie, Université de Krakow, 2016, p. 896-899.
- STOKES, Peter A., MARQUES DE MATOS, Debora et JAKEMAN, Neil, « Analysing Visual Communication in Theory and Practice with Archetype », in *Digital Humanities 2019 Book of Abstracts*, Utrecht, 2019. URL : <https://dev.clariah.nl/files/dh2019/boa/1059.html>
- THORN, Caroline and THORN, Frank, *Domesday Book*, vol. 8-10, Chichester, Phillimore, 1979-1985.
- TREHARNE, Elaine, « The Dates and Origins of Three Twelfth-Century Manuscripts », in *Anglo-Saxon Manuscripts and their Heritage*, Philip PULSIANO et Elaine TREHARNE (éd.), Aldershot, Ashgate, 1998, p. 227-252.
- TREHARNE, Elaine, « The Production and Script of Manuscripts Containing English Religious Texts in the First Half of the Twelfth Century », in *Rewriting Old English in the Twelfth Century*, Mary SWAN et Elaine TREHARNE (éd.), Cambridge, Cambridge University Press, 2000 (Cambridge studies in Anglo-Saxon England, 30), p. 11-40.
- TREHARNE, Elaine, « Cambridge, Corpus Christi College, MS 367 », in *The Production and Use of English Manuscripts 1060-1220*, Orietta DA ROLD, Takako KATO, Mary SWAN et Elaine TREHARNE (éd.), Leicester, University of Leicester, 2010. URL : <https://www.le.ac.uk/english/em1060to1220/mss/EM.CCCC.367.htm>