Les notices d'autorité en XML-TEI : un outil pour l'accroissement collaboratif de connaissances et l'indexation d'éditions de sources

*Authority records in XML-TEI: a tool for collaborative knowledge building and
indexing of source editions*

*Le voce di autorità in XML-TEI: uno strumento per la costruzione di conoscenza
collaborativa e l'indicizzazione delle edizioni di fonte*

Marie Bisson et Anne Goloubkoff

Édition électronique

URL : <http://journals.openedition.org/tabularia/4176>
ISSN : 1630-7364

Éditeur :

CRAHAM - Centre Michel de Boüard, Presses universitaires de Caen

Référence électronique

Marie Bisson et Anne Goloubkoff, « Les notices d'autorité en XML-TEI : un outil pour l'accroissement collaboratif de connaissances et l'indexation d'éditions de sources », *Tabularia* [En ligne], Les sources des mondes normands à l'heure du numérique, mis en ligne le 03 mars 2020, consulté le 04 mars 2020. URL : <http://journals.openedition.org/tabularia/4176>

Les notices d'autorité en XML-TEI: un outil pour l'accroissement collaboratif de connaissances et l'indexation d'éditions de sources

Authority records in XML-TEI: a tool for collaborative knowledge building and indexing of source editions

Le voce di autorità in XML-TEI: uno strumento per la costruzione di conoscenza collaborativa e l'indicizzazione delle edizioni di fonte

Marie BISSON

Pôle Document numérique – MRSH (USR 3486),
Crahaam-Centre Michel de Bouïard (UMR 6273),
Université de Caen Normandie

Anne GOLOUBKOFF

Pôle Document numérique – MRSH (USR 3486),
Université de Caen Normandie

Résumé:

Cet article présente un outil conçu au pôle Document numérique de la Maison de la Recherche en Sciences Humaines de l'université de Caen Normandie. L'enjeu de cet outil est de permettre d'une part la création et l'enrichissement cumulatif d'un *thesaurus* (une base de données XML) à partir des informations collectées dans les textes et d'autre part une indexation fine des textes. Cet environnement de travail est collaboratif: il permet à plusieurs chercheurs travaillant sur différentes sources d'enrichir des bases de données XML partagées. Le contexte dans lequel cet outil a été créé est celui de l'édition de sources anciennes, c'est-à-dire de l'édition scientifique avec des objectifs de recherche et d'analyse des données, mais sans jamais perdre de vue les objectifs d'édition matérielle, soit la possibilité de publier les résultats.

Mots-clés: *Thesauri*, autorités, XML-TEI, index, édition critique

Abstract:

This article presents a tool designed at the pôle Document numérique of the Maison de la Recherche en Sciences Humaines of the University of Caen Normandie. Its goal is, on one hand, to foster the creation and cross enrichment of thesaurus (an XML database) based on information collected through texts and, on the other hand, to provide a fine indexing of the information found in texts. This tool is collaborative: it allows any researcher working on different editions of sources to enrich a shared XML database. The context of inception of this tool was about publishing old sources, meaning scientific publishing with research and data analysis objectives, but still considering the objectives of material publishing, with the capability of publishing the output of the research work.

Keywords: *Thesauri*, authority files, XML-TEI, indices, critical edition

Tabularia « Études », « Les sources des mondes normands... », 2020, p. 1-18, 3 mars 2020

URL: <http://journals.openedition.org/tabularia/4176> | DOI: 10.4000/tabularia.4176

Riassunto:

Questo articolo presenta uno strumento concepito presso il pôle Document numérique della Maison de la Recherche en Sciences Humaines dell'Università di Caen Normandie. Il suo obiettivo è, da un lato, quello di favorire la creazione e l'arricchimento incrociato del thesaurus (un database XML) sulla base delle informazioni raccolte attraverso i testi e, dall'altro, quello di fornire un'accurata indicizzazione delle informazioni contenute nei testi. Questo strumento è collaborativo: permette a qualsiasi ricercatore che lavori su diverse edizioni di fonti di arricchire dei database XML condivisi. Il contesto di partenza di questo strumento è stato quello della pubblicazione di fonti antiche, ovvero l'editoria scientifica con obiettivi di ricerca e di analisi dei dati, ma sempre tenendo conto degli obiettivi della pubblicazione di materiale, con la possibilità di pubblicare l'output del lavoro di ricerca.

Parole chiave: Thesauri, voce di autorità, XML-TEI, indici, edizione critica

L'outil présenté ici a été conçu au pôle Document numérique de la Maison de la Recherche en Sciences Humaines de l'université de Caen Normandie. Il s'agit d'un écosystème permettant la création de *thesauri*, l'indexation des sources éditées numériquement et leur visualisation. Son enjeu est de permettre un enrichissement cumulatif de bases de données XML-TEI à partir d'informations collectées dans les textes et de permettre une indexation fine de ces informations. Cet environnement de travail est collaboratif: il permet à plusieurs chercheurs travaillant sur différentes éditions de sources d'enrichir des bases de données XML-TEI partagées. Le contexte dans lequel cet outil a été créé est celui de l'édition de sources anciennes, c'est-à-dire de l'édition scientifique avec des objectifs de recherche et d'analyse des données, mais sans jamais perdre de vue les objectifs d'édition matérielle: la possibilité de publier les résultats.

Contexte

L'édition scientifique suit des usages bien établis pour l'indexation¹. Selon les périodes et les disciplines, on trouve des index *nominum*, index *locorum*, index *rerum* dont la présentation est généralement normée². Traditionnellement, l'identification et l'indexation consistaient en deux étapes séparées, un héritage de l'époque où le livre était imprimé au plomb: l'identification était réalisée lors de l'annotation du texte avant l'indexation réalisée une fois la mise en page du texte stabilisée et l'édition matérielle bien avancée. Pour des monographies en

1. Concernant l'histoire de l'indexation, voir MANIEZ et MANIEZ, 2009.

2. Dans ses *Conseils pour l'édition des textes médiévaux*, le comité des travaux historiques et scientifiques de l'École des Chartes consacre une partie de son fascicule *Actes et Documents d'archives aux index* (GROUPE DE RECHERCHES « LA CIVILISATION DE L'ÉCRIT AU MOYEN ÂGE », 2001, p. 159-171). Dans le fascicule III consacré aux textes littéraires, la partie dédiée aux index renvoie au fascicule II pour les index de personnes et de lieux mais ajoute des recommandations et des types d'index (GROUPE DE RECHERCHES « LA CIVILISATION DE L'ÉCRIT AU MOYEN ÂGE », BOURGAIN et VIELLIARD, 2002, p. 97-100). Les Sources chrétiennes donnent également des recommandations dans leurs directives pour la préparation des manuscrits (SOURCES CHRÉTIENNES, 2001, p. 81-86).

plusieurs volumes ou des collections, il n'était pas rare de voir un volume propre consacré aux index³. Les logiciels de traitement de texte ont ensuite permis à l'éditeur scientifique de créer des index avant le dépôt du manuscrit auprès de l'éditeur matériel. Les deux étapes se sont alors rejointes : l'éditeur scientifique choisit une forme normalisée et pose un marqueur devant chaque forme de la source qu'il souhaite indexer. Cette méthode à base de marqueur est également celle appliquée par les « logiciels-métier » d'édition matérielle tel qu'Adobe InDesign, prévu pour la mise en page.

Depuis plusieurs années, le pôle Document numérique travaille à la création d'environnements pour l'édition scientifique de sources anciennes en XML-TEI. Il les conçoit⁴ en prenant en compte les besoins de structuration des éditeurs scientifiques (objectifs de recherche) sans perdre de vue les restitutions matérielles de cet encodage (objectifs de publication), par exemple pour l'édition d'inventaires anciens⁵ qui permet d'encoder des listes de livres en XML-TEI (fig. 1).

Fig. 1: Environnement de travail pour l'édition d'inventaires anciens en XML-TEI

3. Voir par l'exemple l'édition du *Magnum theatrum vitae humanae* de Laurens Beyerlinck de 1631 en 8 tomes. Les sept premiers sont consacrés à l'œuvre elle-même, le huitième est un volume d'index. Voir notamment les exemplaires de l'ancienne bibliothèque du Mont Saint-Michel: Avranches, Bibliothèque patrimoniale, A 152-A 159 : https://www.unicaen.fr/bvmsm/ead.html?id=FR_UCBN_MSM_impr_av&c=FR_UCBN_MSM_impr_av_A152-A159.
4. Les environnements de travail sont décrits et téléchargeables sur la page Outils du site du pôle Document numérique : http://www.unicaen.fr/recherche/mrsh/document_numerique.
5. L'environnement TEI_inventairesAnciens a été conçu dans le cadre de l'Equipex Bibliissima afin d'éditer scientifiquement et numériquement des listes de livres. Il est librement téléchargeable sur le site du pôle Document numérique à l'adresse suivante : http://www.unicaen.fr/recherche/mrsh/document_numerique/outils/inventairesAnciens. Le respect du schéma fourni dans cet environnement est une des conditions pour publier un ouvrage au sein de la collection Thecae des Presses universitaires de Caen, en ligne à l'adresse suivante : <https://www.unicaen.fr/services/puc/sources/thecae/>.

Dans les recommandations de la *Text Encoding Initiative*⁶, la solution « classique » d'indexation est l'apposition d'un élément `<index>` qualifié par un attribut `@indexName` précisant la typologie de l'index. À l'intérieur de cet élément, se trouve un élément `<term>` avec la forme normalisée retenue comme point d'accès de l'index. Il s'agit d'une méthode par ancre. On place devant ou derrière chaque concept indexé sa forme normalisée. Cette méthode convient parfaitement pour la publication papier, dans laquelle il s'agit de proposer une liste de formes normalisées renvoyant aux pages des occurrences du texte. Elle est aussi applicable à l'édition électronique : il est possible de construire des index et de renvoyer vers une page HTML et même plus précisément vers le vers ou le paragraphe où se trouve l'occurrence. C'est ainsi que le glossaire de l'édition multisupport du *Roman du Mont Saint-Michel*⁷ a été réalisé (fig. 2 et 3).

Accueil | Sommaire | Chapitre | Glossaire | le *Roman*
du Mont Saint-Michel

Présentation de l'œuvre | Conventions d'édition |

Glossaire

Accès à la [présentation et aux abréviations du glossaire](#)

A B C D E F G H I J K L M N O P

a: excl. 296: *ab!*

[abandonner]: v. tr. P.P. m.sg. abandonné 3710: *laissé à l'abandon*, abandonnez 3025: *grand ouvert*

[abassier]: v. tr. P.P. f.sg. abassiee 1421: *affaiblir*

abeiceis: s.m. 831: *alphabet* («a, b, c»)

abes: s.m.sg. 587, 635, 669, 676, 686, 2105, 2129, 2132, 2153, 2308, 2343, 2345, 2819, 2909, 2925, 2978, 2997, 3269, 3815, 4059, 4069, 4097; abbes 2443, 2449, 2881; abei sg. 596, 615, 653, 2113, 2367, 3363; pl. 890, 909, 1083, 2328; abé sg. 604, 2091, 2128, 2190, 2294, 2298, 3258, 4053; pl. 2354; abbé sg. 2413, 2879: *abbé*

acceptables: adj. 3663: *agréable*

achaison: s.f. 128; acheison 3443: *cause, motif*; acheison 1928, 3300: *circonstances*; par achaisun 1031: *le cas échéant*

Fig. 2: Page de glossaire de l'édition numérique du *Roman du Mont Saint-Michel*

6. Recommandations en ligne à l'adresse suivante: <http://www.tei-c.org/>.

7. GUILLAUME DE SAINT-PAIR, *Le Roman du Mont Saint-Michel (XII^e siècle)*, C. Bougy (éd.), en ligne à l'adresse suivante: <https://www.unicaen.fr/puc/sources/gsp/accueil>.

Fig. 3: Page de résultats depuis l'entrée du glossaire « acceptables » dans le texte du *Roman du Mont Saint-Michel*, livre III, v. 3663. Le vers où l'occurrence recherchée apparaît est mis en surbrillance

La même méthode est appliquée dans l'édition bilingue du *De piscibus*⁸ dont l'index des noms de poissons latins renvoie vers les chapitres mentionnant le nom de l'animal (fig. 4). Les entrées d'index sont alors encodées ainsi :

Ex.: [...] Plinius libro IX. <index indexName="piscis"><term>anguilla</term></index>Anguillae octonis vivunt annis [...]

Exemple tiré de l'encodage du *De piscibus*, ch. 2, paragr. 2.

8. JACQUEMARD, GAUVIN et LUCAS-AVENEL (éd.), 2013, en ligne à l'adresse suivante : <https://www.unicaen.fr/puc/sources/depiscibus/>.

Fig. 4 : Page de l'index des noms latins de poissons de l'édition du *De piscibus*

Il y a cependant une limite à cette méthode: puisqu'un marqueur est posé devant la forme indexée, la forme originale ne peut pas être marquée (mise en surbrillance par exemple) lors du retour au texte, alors que c'est un usage du web auquel le lecteur peut légitimement s'attendre. L'éditeur peut tout au plus renvoyer à l'« élément ancêtre » commun à toutes les occurrences, généralement le vers <l> ou le paragraphe <p>. Enfin, il n'y a aucune autre exploitation possible que de pointer vers des occurrences de termes. Or, dans les textes anciens, l'éditeur scientifique se trouve face à une grande variété de formes qu'il peut avoir envie de récupérer, de recenser, d'analyser.

Le vocabulaire de la TEI permet par ailleurs d'encoder les anthroponymes; nous pourrions nous arrêter à cette variante consistant à l'encadrement de l'entrée d'index et de l'occurrence au sein d'un élément <persName>.

Ex.: [...] Saint <persName><index indexName="anthroponyme"><term>Aubert (saint) / Aubertus</term></index>Aubert</persName>, 12<hi rend="sup">e</hi> evesque d'Avranches, fit bastir la premiere <lb/>église [...]

Exemple forgé à partir du texte de la *Bresve histoire de l'abbaye du Mont Saint-Michel* par dom Thomas le Roy⁹, ch. 1.

Cependant une fois le concept encodé, l'annotateur souhaiterait y ajouter les informations qui l'ont amené à cette identification : des dates de naissance, de décès, etc. ; il peut être tenté également de récapituler les informations qu'il trouve dans son texte ou dans sa bibliographie. Tel est l'objet des notices d'autorités institutionnelles, notamment celles de la BnF¹⁰ ou d'Idref¹¹. Nous avons essayé d'utiliser ces notices d'autorité. Si pour les lieux géographiques et des auteurs d'ouvrage, cela pouvait s'avérer cohérent et utilisable, nous étions cependant confrontés à plusieurs limites. L'éditeur scientifique ne peut pas modifier les notices (dont le protocole de modification est très rigoureux et réservé aux membres des institutions) ; il ne peut pas créer ni extraire de sous-entrées ; enfin, les seules bases existantes sont les lieux, les œuvres et les personnes (celles ayant participé au processus de création d'un document : auteur, copiste, enlumineur, dessinateur, imprimeur, etc.). Or, nous voulions pouvoir créer des index thématiques (voir notamment l'exemple des poissons *infra*) et indexer tous les noms de personnes que nous rencontrions – même les plus obscurs, rencontrés dans un seul document. Nous avons donc écarté l'idée d'utiliser les référentiels d'autorités nationaux ou internationaux directement.

À partir des recommandations de la TEI et des besoins pour l'indexation – papier et électronique –, nous avons décidé d'adapter dans certains cas¹² la méthode traditionnelle décrite précédemment. Nous avons donc identifié nos besoins et nos exigences :

- marquer exclusivement la forme d'origine lors du retour au texte dans une édition en ligne ;
- pouvoir recenser les formes rencontrées (formes principales et variantes) ;
- normaliser des formes plus facilement à l'échelle d'une collection, d'un laboratoire, etc. ;
- pointer vers des référentiels nationaux ou internationaux quand ils existent (BnF, Viaf, etc.) ;
- permettre l'indexation hiérarchique (entrées et sous-entrées) ;
- accumuler de la connaissance collectivement.

9. Texte édité dans le cadre de la thèse de Marie Bisson (voir BISSON, 2015 et plus spécifiquement t. 2, p. 6).

10. <https://data.bnf.fr/fr/>.

11. <https://data.idref.fr/>.

12. Ce travail a été entrepris au départ pour l'édition de deux corpus : le corpus de dom Thomas Le Roy consacré à l'histoire du Mont Saint-Michel (voir BISSON, 2015) et le corpus des actes des évêques d'Évreux (*Les actes des évêques d'Évreux (XI^e siècle-1223)*, G. Combalbert (éd.), URL : <https://www.unicaen.fr/puc/sources/ecartae/evreux/>).

Actuellement, quatre bases coexistent. Elles sont partagées et ouvertes à une communauté plus ou moins restreinte selon les cas. Elles présentent des caractéristiques communes et d'autres plus spécifiques :

- *thesaurus* des noms de personnes ;
- *thesaurus* des noms de lieux ;
- *thesaurus* des titres d'œuvres ;
- *thesaurus* des noms de poissons et créatures aquatiques trouvés dans les traités latins d'ichtyologie.

Ces bases permettent l'indexation des formes mais aussi l'accumulation des informations de façon partagée. Avant de détailler la réalisation technique, voici un état des lieux des bases telles qu'elles se présentent à ce jour : chaque partie consacrée à une base récapitule les besoins auxquels nous avons dû répondre.

État des lieux des bases existantes

Thesaurus Personnes

En octobre 2019, le *thesaurus* des noms de personnes comprenait 11 250 notices utilisées dans le cadre de plusieurs programmes de recherche – Thecae¹³, Actépi¹⁴, *e*-Cartae¹⁵, Vexicaen¹⁶, Normonde¹⁷, *Ex Monasterio Montis Sancti Michaelis*¹⁸, Ichtya¹⁹ – par des chercheurs de plus de six institutions différentes.

13. La collection Thecae, publiée par les Presses universitaires de Caen (PUC), a été définie et mise en œuvre par l'Equipex Bibliissima. Elle accueille des éditions d'inventaires médiévaux et modernes, du VIII^e au XVIII^e siècle, et est accessible à l'adresse suivante : <https://www.unicaen.fr/services/puc/sources/thecae/>.
14. Actépi est une ANR portée par Grégory Combalbert (Craham, UMR 6273), intitulée *Actes épiscopaux français du Moyen Âge: édition multimodale et exploitation (2019-2022)*. Elle est construite en partenariat avec les équipes Polen (EA 4710, Orléans), Crulh (EA 3945, Lorraine), Arche (EA 3400, Strasbourg) et Tempora (EA 7468, Rennes).
15. *E*-Cartae est un outil d'édition critique, de publication et de consultation des corpus de chartes médiévales, développé par G. Combalbert et le pôle Document numérique de la MRSN. Sous sa forme actuelle, *e*-Cartae propose l'édition critique du corpus des actes des évêques d'Évreux, du XI^e siècle à 1223.
16. Vexicaen est un programme de recherche (Feder) dirigé par Pierre Bauduin (Craham, UMR 6273) et décrit à l'adresse suivante : <http://www.unicaen.fr/craham/spip.php?article1208>. Il a notamment porté sur l'expérimentation de l'indexation des noms de personnes et de lieux de chartes du corpus *Scripta*.
17. Dans le cadre du projet RIN Normonde (resp. J. F. Klein) la constitution d'une prosopographie des Normands a été prévue par P. Bauduin et A. Peters-Custot afin notamment d'établir leurs déplacements et influences géographiquement (notamment entre l'Angleterre, la France et l'Italie du Sud).
18. Ce programme de recherche, dirigé par Catherine Jacquemard (Craham, UMR 6273), a pour objectif d'étudier la bibliothèque monastique du Mont Saint-Michel. Le *thesaurus* Personnes a été utilisé afin d'indexer toutes les autorités mentionnées dans les sources publiées et les catalogues des ouvrages.
19. Ce programme de recherche du Craham s'intéresse à la transmission du savoir ichtyologique (resp. T. Buquet et P. Bauduin). Le *thesaurus* Personnes est utilisé pour l'indexation des sources éditées et des catalogues.

Cette base soutenue par l'Equipex Biblissima²⁰ et le consortium Cosme²¹ est maintenue par le pôle Document numérique.

Une fiche Personne contient *a minima* : un identifiant ; le nom du créateur de la notice ; une forme nominale en français ou en latin ; les dates de naissance et de décès ; le sexe. L'encodeur peut y ajouter de façon normalisée :

- une ou plusieurs formes normalisées dans des langues différentes et des variantes ;
- des informations relatives à la naissance et au décès (dates et lieux) ;
- des éléments d'identification indexés au moyen d'un index des Occupations. Les lieux relatifs à ces identifications peuvent eux-mêmes être liés au *thesaurus* des noms de lieux (cf. description *infra*) ;
- des lieux de résidence, des lieux traversés ;
- de la bibliographie générale ou relative à chacune des informations données ;
- des liens hypertextes ;
- des liens vers des bases d'autorité de référence²² ;
- des attestations dans des documents ;
- des notes de relations.

L'ensemble des noms de personnes, des occupations et des lieux cités peuvent / doivent être indexés au moyen des bases du même nom, c'est-à-dire que ces données (en vert sur la figure 5) sont liées grâce à un identifiant. Par ailleurs, la granularité des données peut être enrichie pour des projets spécifiques, notamment prosopographiques.

Actuellement, la base contient principalement des personnages ayant un lien avec la Normandie. Cela s'explique par les corpus dépouillés : actes des évêques d'Évreux, corpus historiographique consacré au Mont Saint-Michel. Mais elle devrait s'hétérogénéiser avec le dépouillement des corpus d'inventaires anciens de livres²³ et des corpus de chartes d'évêques²⁴.

L'encodage XML-TEI complet de la notice d'Achard de Saint-Victor est disponible dans le *Manuel d'encodage d'une fiche Personne*²⁵.

20. <https://projet.biblissima.fr/>.

21. <https://cosme.hypotheses.org/>.

22. Il s'agit de renvoyer au référentiel des autorités de la BnF et au fichier d'autorité international virtuel, Viaf (<https://viaf.org/>).

23. Citons notamment le projet de publication des inventaires des manuscrits de Gembloux (SHAPOVALOVA, à paraître) ou celui des inventaires des livres provenant de l'abbaye du Mont Saint-Michel (Bisson, à paraître).

24. Projet Actépi décrit *supra*.

25. Voir BISSON, 2019, p. 24.

<p>Achard de Saint-Victor <i>Achardus Sancti Victoris</i> ¹ Achard ² Maître Achard Naissance : 1100? Mort : le 29 ou 30 mars 1170 ³ Éléments d'identification Abbé de Saint-Victor de Paris de 1153 à 1161 ⁴ Evêque d'Avranches ⁵ de 1161 à v. 1170 Attestation(s) ...1162-02-25... ⁶ Notes 1. SPEAR 2006, p. 5. 2. T.L.R., <i>HA</i>, ch. 16 § 15. 3. SPEAR 2006, p. 5. 4. 1155-1161 selon Spear (SPEAR 2006, p. 5). 5. T.L.R., <i>CR</i>, ch. 16 § 11 et ch. 42 § 190. 6. SPEAR 2006, p. 5.</p>	<p>Source : OPP N° de notice : PDDN_P.1389343678086 Lien vers les autorités BNF : la fiche autorité. Lien vers les autorités VIAF : la fiche autorité. Créateur(s) de la notice : MARIE BISSON</p>
--	---

Fig. 5: Exemple d'affichage d'une notice Personne sur le site web de visualisation

<p>Avranches <i>Abrincae</i> <i>Abrincatensis</i> Éléments de localisation Pays : France Département : Manche Latitude : 48.68823709322551 Longitude : -1.3564802368162439 Sous-entrées Sous-entrée n° 1 : Saint-André Description : église Sous-entrée n° 2 : Diocèse d'Avranches Description : diocèse</p>	<p>N° de notice : PDDN_L.2 Créateur(s) de la notice : MARIE BISSON CLÉMENTINE BERTHELOT GREGORY COMBALBERT</p>
--	--

Fig. 6: Exemple d'affichage d'une notice Lieu sur le site web de visualisation

Thesaurus *Lieux*

En octobre 2019, le *thesaurus* des noms de lieux comptait 5 752 notices utilisées dans le cadre de plusieurs programmes de recherche – Thecae, Actépi, e-Cartae, Vexicaen, Normonde, *Ex Monasterio Montis Sancti Michaelis*, Ichtya – par des chercheurs de plus de six institutions différentes. 4 018 de ces notices sont géolocalisées (centroïde des communes françaises d'avant 2015). Cette base soutenue par l'Equipex Biblissima et le consortium Cosme est maintenue par le pôle Document numérique.

Une fiche Lieu contient *a minima* : un identifiant ; un créateur ; une forme nominale en français ou en latin ; la qualification du lieu selon une liste fermée ; des éléments de localisation. Elle peut en outre être complétée de variantes de noms et de sous-entrées. Ces sous-entrées correspondent à des institutions clairement rattachées au lieu (fig. 6).

En décembre 2017, une chaîne de traitement a été créée au pôle Document numérique permettant de générer pour les lieux français :

- la géolocalisation (centroïde des communes d'avant 2015) ;
- la mise à jour des cantons, communes nouvelles.

Cela permet de positionner chacun des lieux présents dans la base sur une carte, mais surtout de géolocaliser les lieux indexés au sein des sources éditées utilisant ces ressources.

Thesaurus *Œuvres*

En octobre 2019, le *thesaurus* des titres d'œuvres comptait 1 888 notices utilisées dans le cadre de plusieurs programmes de recherche – Thecae, *Ex Monasterio Montis Sancti Michaelis*. Cette base soutenue par l'Equipex Biblissima est maintenue par le pôle Document numérique. Une notice *a minima* contient un identifiant ; un créateur de notice ; un auteur – lié à la base Personnes (en bleu sur la figure 7) ou mentionné comme anonyme – ; un titre retenu dans la langue originale du texte. La fiche peut également indiquer un lien vers des référentiels d'autorités nationaux ou internationaux ; des variantes de titres ; une date ; des notes de commentaire et une bibliographie de référence.

Ce *thesaurus* a été créé pour permettre l'indexation des œuvres des éditions d'inventaires anciens, mais également pour établir une bibliographie de références aux œuvres identifiées dans des catalogues de livres (par exemple dans le cadre de la *Bibliothèque virtuelle du Mont Saint-Michel*²⁶).

26. Voir *Bibliothèque virtuelle du Mont Saint-Michel*. URL : <https://www.unicaen.fr/bvmsm/>.

Auteur : Alcuin	Source :
Titre retenu : <i>Compendium in Canticum canticorum</i>	n° de notice : PDDN_w.259
Date : 07..?	Lien vers les autorités BNF : la fiche autorité.
Commentaire : Commentaire exégétique et allégorique du Cantique des cantiques.	Créateur(s) de la notice : MARIE BISSON
Bibliographie de référence	
ALCUIN ; FOLIOT Gilbert, <i>Gilberti Foliot Episcopi Londinensis, Expositio in Canticum canticorum, una cum Compendio Alcuini. Nunc primum à Bibliotheca Regia in lucem prodit, opera & studio Patricii Junii Bibliothecarii Regii</i> , Londini [Londres], ex typographio Regio [Robert Barker], MDCCXXXVIII [1638].	
ALCUIN, <i>Commento al Cantico dei Cantici: con i commenti anonimi "Vox Ecclesie", "Vox antique Ecclesie"</i> , éd. Rossana E. Guglielmetti, Firenze, SISMEL Ed. del Galluzzo (Millennio Medievale. Testi ; 13), 2004.	

Fig. 7 : Exemple d'affichage d'une notice Œuvre sur le site web de visualisation

<i>Albirem</i>	n° de notice :
<i>Albyroz</i>	PDDN_PISC.1458913661693
Identifications répertoriées	
Identification 1 : l'éponge.	
Étude(s) : Jacquemard, Gauvin & Avenel 2013, 92-93.	
Source(s) : HS, 4, 1.	
Commentaire : Il faut reconnaître dans l' <i>albirem</i> / <i>albyroz</i> une sorte d'éponge, réputée dans l'Antiquité gréco-romaine pour la qualité de son tissu et désignée sous le nom d' <i>achilleios</i> / <i>achillium</i> (voir Stadler 1920, 1604, et Kitchell & Resnick 1999, 1662, n. 45 ; voir aussi D'Arcy Thompson 1947, 23, pour l' <i>achilleios</i> grec ; De Saint-Denis 1947, 1, pour l' <i>achillium</i> latin). La description de l' <i>albirem</i> / <i>albyroz</i> chez les encyclopédistes médiévaux (AM, DA, 24, 6 (9) ; TC, nat., 7, 6 ; et VB, SN, 17, 29) remonte à ARIST. HA 548 b 2 MS.	
Relation(s) : Voir aussi <i>Achillium</i> ; <i>Nubes</i> ; <i>Sfungia</i> ; <i>Spongia</i> ; <i>Tragus</i> .	

Fig. 8 : Exemple d'affichage d'une notice Poisson sur le site web de visualisation

Thesaurus des noms de poissons et des créatures aquatiques trouvés dans les corpus latins

Le *thesaurus* des noms de poissons et de créatures aquatiques a été créé dans le cadre du programme de recherche Ichtya ; son utilisation et son enrichissement sont actuellement restreints aux membres de l'équipe du même nom. La base est étroitement liée à la Bibliothèque Ichtya²⁷, un corpus de textes latins,

27. Voir *Bibliothèque Ichtya: corpus de traités latins d'ichtyologie*, à paraître en ligne.

antiques et médiévaux, consacrés à l'ichtyologie. Au 6 octobre 2019, la base comptait 1 610 notices.

Une notice contient *a minima* : un identifiant ; un créateur de notice ; un nom de poisson / créature aquatique latin ; une identification (une forme française ; une éventuelle forme scientifique ; l'étude ayant fait l'identification ; une source mentionnant la forme). Elle peut être enrichie d'un lien vers le *thesaurus* opentheso Zoomathia²⁸ ; des variantes de noms ; de plusieurs identifications ; de notes de commentaire sur l'identification ; d'un commentaire général sur la notice ; des notes de relation entre poissons (un lien pointe alors vers la notice du poisson concerné) (fig. 8).

Réalisation technique

La solution technique que nous proposons sépare les sources, éditées par un chercheur ou une équipe de chercheurs, des ressources qui vont être partagées par une communauté (un programme de recherche, un ou plusieurs laboratoires, des éditeurs scientifiques au sein d'une collection, etc.). Chaque entrée dispose de sa propre notice XML-TEI créée dans une base XML thématique – actuellement Personnes, Lieux, Œuvres, ou plus thématiquement les noms latins de Poissons ou créatures aquatiques (fig. 9).

Fig. 9 : Schéma représentant les bases XML Sources et ressources

28. Zoomathia est un RI porté par le Cepam (UMR 6472) dirigé par Arnaud Zucker, consacré à la transmission culturelle des savoirs zoologiques (Antiquité-Moyen Âge), qui rassemble neuf équipes internationales. Opentheso est un gestionnaire de *thesaurus* multilingue, développé par la plateforme technologique TRD (Têtes des Réseaux Documentaires) située à la Maison de l'Orient et de la Méditerranée en partenariat avec le GDS-FRANTIQ.

Le lien entre la forme rencontrée dans la source et sa notice parmi les ressources se fait par le moyen d'un identifiant. Les ressources, pour être accessibles aux différents éditeurs scientifiques qui en ont besoin, sont hébergées sur un serveur. L'éditeur scientifique qui accepte de participer à la constitution du thesaurus va pouvoir, quand il identifie un anthroponyme, un toponyme, un titre d'œuvre ou un nom de poisson :

- utiliser les notices déjà incrémentées par la communauté d'éditeurs participants ;
- créer une notice ;
- modifier une notice pour l'améliorer ou la corriger.

Pour permettre ce travail, nous avons conçu un écosystème s'appuyant sur :

- un logiciel ;
- un plug-in collaboratif ;
- un environnement de saisie de notices ;
- un site autonome de consultation des données.

Les éditeurs de sources anciennes de Caen, en collaboration avec le pôle Document numérique ont l'habitude de travailler avec le logiciel «XMLMind XML Editor», un éditeur XML configurable dont la version personnelle est téléchargeable gratuitement²⁹ et pour lequel le pôle Document numérique a développé plusieurs environnements d'édition ou de catalogage³⁰.

Pour la création des ressources, nous avons donc décidé de garder le même écosystème de travail. Dans le cadre de l'Equipex Biblissima, un plug-in collaboratif (PluCo)³¹ a été conçu pour être utilisé avec le logiciel précédemment décrit. Développé par le Certic de l'université de Caen Normandie, ce plug-in permet de mettre en interaction des documents XML hébergés sur un serveur dans une base de données XML (par exemple BaseX³²) et l'éditeur «XMLMind XML Editor». La base de données peut être modifiée et interrogée par les utilisateurs depuis leur interface traditionnelle de travail. Ce plug-in favorise donc le travail collectif et collaboratif. Par ailleurs, il permet l'indexation des éditions de sources en interrogeant les données hébergées à distance (ressources locales ou ressources nationales) en rapatriant les identifiants et éventuellement des contenus. L'identifiant de la ressource correspondante est utilisé comme valeur de l'attribut @ref des éléments suivants: <name> pour les lieux et les personnes, <title> pour les titres d'œuvre, <term> pour les noms de poissons.

29. Logiciel développé par la société Pixware, disponible à l'adresse suivante: <http://www.xmlmind.com/xmleditor/download.shtml>.

30. Une sélection de ces environnements fait l'objet d'une diffusion libre sur la page Outils du pôle Document numérique: http://www.unicaen.fr/recherche/mrsh/document_numerique/outils.

31. Ce plug-in est décrit et téléchargeable à l'adresse suivante: http://www.unicaen.fr/recherche/mrsh/document_numerique/outils/pluco.

32. Décrit et téléchargeable librement à cette adresse: <http://basex.org/>.

Ex. : *Inrumpunt ergo in secreta ac* <term ref="pddn_pisc.1459242964377.xml#pddn_pisc.1459242964377" type="animal" xml:lang="la">vitulos</term> earum aut fetas vel etiamnum gravidas lancinant morsu incursuque ceu Liburnicarum rostris fodiunt [...]

Exemple tiré de la *Bibliothèque Ichtya*³³, PLIN. nat. 9, 13.

Le pôle Document numérique a donc élaboré un environnement modulaire pour normaliser la saisie des notices ressources. L'environnement permet de traiter les différentes spécificités des ressources indexées. Actuellement, il prend en compte la saisie des données pour une base Personnes, une base Lieux, une base Œuvres et une base de noms latins de Poissons avec des vues et commandes contextuelles. Cela permet notamment de proposer un formulaire vierge *a minima* pour renseigner les données de la façon la plus homogène possible, mais aussi l'ajout de champs en fonction de nouveaux projets ou encore l'ajout de nouveaux modules au sein de l'environnement lors de la création de nouvelles bases de ressources.

Cet environnement est constitué :

- de vues contextuelles configurées au moyen du langage CSS ;
- d'un schéma créé à partir des recommandations de la TEI ;
- d'une barre d'outils générique à l'environnement ;
- de modules génériques et de modules spécifiques permettant d'appeler des commandes en fonction du contexte d'encodage.

Pour normaliser au maximum l'encodage des notices, nous avons également rédigé nos choix au sein de manuels (un par type de ressources) qui sont distribués avec l'environnement sur le site du pôle Document numérique³⁴. Le manuel explique chaque choix d'encodage :

- en précisant le nom et les définitions des éléments utilisés ;
- en décrivant le rendu graphique de chaque élément dans l'environnement ;
- en décrivant les commandes existantes ;
- en établissant un glossaire final des éléments utilisés ;
- en fournissant des exemples de notices.

Enfin pour faciliter la consultation des données, nous fournissons aux utilisateurs l'accès à un site web des bases partagées (actuellement sous mot de passe) qui expose les données.

33. Voir *Bibliothèque Ichtya: corpus de traités latins d'ichtyologie*, à paraître en ligne.

34. L'environnement et sa documentation sont accessibles sur le site du pôle Document numérique à l'adresse suivante : http://www.unicaen.fr/recherche/mrsh/document_numerique/outils/thesauri.

Bilan et perspectives

L'utilisation des bases contenant des ressources partagées pour indexer les personnes, les lieux, les œuvres ou les poissons dans les sources permet la création d'index hiérarchisés pour des éditions électroniques et / ou papier. Des index hiérarchisés ont ainsi été générés pour le corpus des actes des évêques d'Évreux³⁵, l'édition du corpus de dom Thomas Le Roy³⁶, l'édition des inventaires anciens du Mont Saint-Michel³⁷. Toutes ces notices fournissent également une aide à la lecture : elles permettent l'identification des personnes, lieux, œuvres, et noms de poissons latins des sources indexées avec accès à tout ou partie de la notice (cf. Bibliothèque Ichtya ; *e-Cartae* ; édition du corpus de dom Thomas Le Roy ; inventaires anciens du Mont Saint-Michel), puisqu'en cliquant sur la forme indexée, on accède à la notice entière. La géolocalisation des lieux rencontrés dans les sources permet ensuite d'accéder aux documents à partir d'une carte. Cette expérimentation a été menée à bien sur *e-Cartae*.

Les perspectives pour ces bases sont multiples. Tout d'abord leur utilisation et la publication des données peuvent être améliorées en permettant l'indexation des références bibliographiques afin de pouvoir (re)générer la forme de ces références selon chaque projet d'édition (références courtes / références longues ; système auteur-date / système classique, etc.). Une bibliothèque de groupe Zotero³⁸ a donc été créée et son usage va être généralisé à l'ensemble des utilisateurs des *thesauri*. Le système d'indexation (l'établissement du lien entre la référence dans la notice et la référence de la base Zotero) fonctionne également avec PluCo. Une autre perspective est de pouvoir incrémenter la base Personnes en ajoutant des données spécifiques pour constituer des bases prosopographiques. Deux expérimentations sont en cours. La première est réalisée dans le cadre du projet Normonde. Sur les données Personnes en lien avec le projet prosopographique sont ajoutés des mots-clés (<keyword> et <term>) qualifiant le projet. L'objectif est d'extraire ces notices qualifiées afin de réaliser un site propre répondant à une cohérence scientifique. L'autre projet fonctionne dans le sens inverse : une prosopographie des abbés normands du XIII^e siècle est constituée dans une base propre³⁹. On peut également envisager le versement de tout ou partie des données dans la base Personnes à la fin du travail.

La troisième piste concerne la création de nouvelles bases. Dans cette perspective, nous pouvons citer la constitution d'un thesaurus de pièces liturgiques du Mont Saint-Michel par Louis Chevalier⁴⁰ dans le cadre de sa thèse de doctorat, qu'il pourrait ouvrir à une communauté. Citons également le souhait de plusieurs chercheurs de constituer un thesaurus *rerum*. Enfin ces bases devraient permettre

35. Voir *Les actes des évêques d'Évreux (XI^e siècle-1223)*, URL : <https://www.unicaen.fr/puc/sources/ecartae/evreux/>.

36. BISSON, 2015.

37. BISSON, à paraître.

38. <https://www.zotero.org/groups/447270/>.

39. Il s'agit du travail de thèse de Fabien Paquet (PAQUET, 2018).

40. CHEVALIER, 2019.

des études de « réseaux » à l'échelle de corpus (réseaux géographiques ; réseaux de personnes). La prochaine étape pour structurer ce travail va être la constitution d'un comité scientifique chargé de définir les priorités et les orientations à venir de ces *thesauri*.

Bibliographie

Sources et bases de données en ligne

- Bibliothèque Ichtya : corpus de traités latins d'ichtyologie*, B. Gauvin (dir.), Caen, Craham – Pôle Document numérique MRSH, à paraître en ligne.
- Bibliothèque virtuelle du Mont Saint-Michel*, M. Bisson (coord.) et C. Jacquemard (dir.), Caen, Craham – Pôle Document numérique MRSH, 2017-2020. URL : <https://www.unicaen.fr/bvmsm/>.
- GUILLAUME DE SAINT-PAIR, *Le Roman du Mont Saint-Michel (XII^e siècle)*, C. Bougy (éd.), Caen, Presses universitaires de Caen, Scriptorial d'Avranches (*Fontes & Paginae*), 2009. URL : <https://www.unicaen.fr/puc/sources/gsp/accueil>.
- Les actes des évêques d'Évreux (XI^e siècle-1223)*, G. Combalbert (éd.), Caen, Craham – Presses universitaires de Caen (*e-Cartae, Actépi*), 2019-2020. URL : <https://www.unicaen.fr/puc/sources/ecartae/evreux/>.
- Les inventaires des livres de l'ancienne abbaye du Mont Saint-Michel*, M. Bisson (dir.), Caen, Presses universitaires de Caen (*Thecae*), à paraître en ligne.
- Scripta. Base des actes normands médiévaux*, P. Bauduin (dir.), Caen, Craham – Pôle Document numérique MRSH, 2010-2020. URL : <http://www.unicaen.fr/scripta/>.

Études

- BISSON, Marie, *Une édition numérique structurée à l'aide de la Text Encoding Initiative des textes montois de dom Thomas Le Roy : établissement critique des textes, recherches sur les sources, présentation littéraire et historique*, thèse de doctorat de littérature et lettres anciennes, Université de Caen Basse-Normandie, UFR HSS (dir. Catherine Jacquemard et Véronique Gazeau), 2015, 2 vol.
- BISSON, Marie, *Rédiger une notice pour le thesaurus des personnes*, Caen, Pôle Document numérique (Les Manuels du pôle), 2019. URL : http://www.unicaen.fr/recherche/mrsh/sites/default/files/public/document_numerique/manuel_personnes.pdf.
- CHEVALIER, Louis, *Agere et Statuere : étude historique et édition critique et numérique des deux ordinaires liturgiques du Mont Saint-Michel (XIV^e-XV^e siècle)*, thèse de doctorat en Histoire, Université de Caen Normandie, UFR HSS (dir. Véronique Gazeau et C. Jacquemard), 2019, 3 vol.
- GRUPE DE RECHERCHE « LA CIVILISATION DE L'ÉCRIT AU MOYEN ÂGE », *Conseils pour l'édition des textes médiévaux. Fascicule II. Actes et documents d'archives*, Paris, Comité des travaux historiques et scientifiques, École nationale des Chartes (Orientation et méthode), 2001.
- GRUPE DE RECHERCHE « LA CIVILISATION DE L'ÉCRIT AU MOYEN ÂGE », BOURGAIN, Pascale et VIELLIARD, Françoise, *Conseils pour l'édition des textes médiévaux. Fascicule III. Textes littéraires*, Paris, Comité des travaux historiques et scientifiques, École nationale des Chartes (Orientation et méthode), 2002.

- JACQUEMARD, Catherine, GAUVIN Brigitte et LUCAS-AVENEL, Marie-Agnès (éd.), *Hortus Sanitatis, Livre IV, Les Poissons*, Caen, Presses universitaires de Caen (*Fontes & Paginae*), 2013. URL : <https://www.unicaen.fr/puc/sources/depiscibus/>.
- MANIEZ, Jacques et MANIEZ, Dominique, *Concevoir l'index d'un livre : histoire, actualité, perspectives*, Paris, ADBS, 2009.
- PAQUET, Fabien, *Des crosses et des couronnes. Pouvoirs abbaciaux et pouvoirs royaux dans le diocèse de Rouen (fin du XII^e siècle-milieu du XV^e siècle)*, thèse de doctorat en Histoire, Université de Caen Normandie, UFR HSS (dir. Véronique Gazeau), 2018, 2 vol.
- SHAPOVALOVA, Anastasia (éd.), *Les inventaires de l'abbaye de Gembloux*, Caen, Presses universitaires de Caen (Thecae), à paraître.
- SOURCES CHRÉTIENNES, *Directives pour la préparation des manuscrits*, Paris, Les éditions du Cerf, 2001.