

ÉCOLE NATIONALE DES CHARTES
UNIVERSITÉ PARIS, SCIENCES & LETTRES

Léo Trotin

licencié ès lettres
diplômé de master

Mise en relation de deux bases de données de recherche en XML-TEI

L'exemple de ThEMA et de SourcEncyMe

Mémoire pour le diplôme de master
« Technologies numériques appliquées à l'histoire »

2024

Résumé

Ce mémoire explore le processus de connexion entre deux bases de données de recherche en XML-TEI, SourcEncyMe et ThEMA, dans le but d'améliorer leur accessibilité et leur utilité pour la recherche scientifique. Le projet a débuté par une analyse des différentes méthodes de mise en relation, suivie d'une étude approfondie des deux bases pour garantir la qualité des données, déterminer les emplacements appropriés pour les liens dans les fichiers XML, et identifier les informations pertinentes à relier. Ensuite, diverses approches ont été examinées pour identifier les textes à connecter. Finalement, des liens ont été créés de manière semi-automatique entre les deux bases, et une indexation automatique des récits identifiés dans SourcEncyMe a été faite avec ThEMA. Cette interconnexion améliore non seulement l'accès à l'information, mais aussi la collaboration scientifique, valorise les données et ouvre de nouvelles perspectives pour la recherche.

Mots-clés : Interopérabilité ; XML-TEI ; Récits exemplaires ; Encyclopédies ; XQuery ; Python ; Automatisation ; ThEMA ; SourcEncyMe ; Open Data ;

Informations bibliographiques : Léo Trotin, *Mise en relation de deux bases de données de recherche en XML-TEI. L'exemple de ThEMA et de SourcEncyMe*, mémoire de master « Technologies numériques appliquées à l'histoire », dir. Enimie Rouquette, École nationale des chartes, 2024.

Remerciements

JE tiens à exprimer ma reconnaissance à ma directrice de mémoire, Enimie Rouquette, pour son encadrement et ses conseils tout au long de ce travail. Je remercie également Jean-Damien Genero, Jean-Paul Rehr et Emmanuelle Kuhry pour leur soutien en tant que tuteurs techniques. Je suis reconnaissant envers Marie-Anne Polo de Beaulieu et Isabelle Draelants pour leur direction scientifique. Mes remerciements vont aussi à l'équipe de l'EHESS et de l'IRHT pour leur accompagnement durant ce stage. Enfin, je remercie le LabEx HaStec pour le financement de ce projet.

Bibliographie

- BEAUGUITTE (Laurent), « L'analyse de réseaux en sciences sociales et en histoire », dans *Le réseau. Usages d'une notion polysémique en sciences humaines et sociales*, dir. Rosemonde Letricot, Mario Cuxac, Maria Uzcategui Moncada et Andréa Cavaletto, Louvain-La-Neuve, 2016, p. 9-24, URL : <https://shs.hal.science/halshs-01476090> (visité le 31/08/2024).
- BERLIOZ (Jacques) et BEAULIEU (Marie Anne Polo de), « Introduction générale », dans *Le tonnerre des exemples : Exempla et médiation culturelle dans l'Occident médiéval*, dir. Marie-Anne Polo De Beaulieu et Pascal Collomb, Rennes, 2010, p. 11-15, DOI : 10.4000/books.pur.132003.
- BERLIOZ (Jacques) et POLO DE BEAULIEU (Marie-Anne), « Les recueils d'exempla et la diffusion de l'encyclopedisme médiéval », dans *L'enciclopedismo medievale*, dir. Michelangelo Picone, Ravenna, 1994, p. 179-212.
- BREMOND CLAUDE, SCHMITT JEAN-CLAUDE HISTORIEN et LE GOFF JACQUES, *L'Exemplum*, Turnhout, 1982.
- BURGHART (Marjorie), « L'informatisation du ThEMA », *Les Cahiers du Centre de Recherches Historiques*, 35 (2005), DOI : 10.4000/ccrh.3025.
- BURNARD (Lou), *Qu'est-ce que la Text Encoding Initiative ?*, Marseille, 2015, URL : <https://books.openedition.org/oep/1237> (visité le 11/04/2024).
- DRAELANTS (Isabelle) et KUHRY (Emmanuelle), « Encyclopédies médiévales en milieu numérique : Les nouveaux enjeux de SourcEncyMe pour le traitement des auctoritates », dans *Dix ans de corpus d'auteurs*, dir. Fatiha Idmhand et Ioana Marasescu-Galleron, éditions des archives contemporaines, Paris, 2022, p. 155-178, DOI : 10.17184/eac.5479.
- HOLTZ (Louis), « L'institut de recherche et d'histoire des textes (IRHT) », *Les Cahiers du Centre de Recherches Historiques*, 36 (2005), DOI : 10.4000/ccrh.3046.
— *L'IRHT, au fil des ans*, 2015, DOI : 10.58079/qhxe.
- JIANG (Tianyin), HOU (Yixin) et YANG (Jaebum), « Literature Review on the Development of Visualization Studies (2012–2022) », *Engineering Proceedings*, 38–1 (2023), p. 1-10, DOI : 10.3390/engproc2023038089.
- KUHRY (Emmanuelle), *Quels traitements numériques pour les corpus encyclopédiques ? Les projets SourcEncyMe et Mythologia en dialogue (retour sur la séance du séminaire*

- « *Les p'tits déj' 'Humanités numériques' de l'IRHT* », 14/01/22), 2022, DOI : 10.58079/nmg9.
- LE GOFF (Jacques), « Pourquoi le XIII^e siècle a-t-il été plus particulièrement un siècle d'encyclopedisme ? », dans *L'enciclopedismo medievale*, dir. Michelangelo Picone, Ravenna, 1994, p. 23-40.
- NURMIKKO-FULLER (Terhi), *Linked Data for Digital Humanities*, London, 2023.
- OLDONI (Massimo), « Giovanni da San Gimignano », dans *L'enciclopedismo medievale*, dir. Michelangelo Picone, Ravenna, 1994, p. 213-228.
- PEDRO (Bruno), *Building an API Product : Design, Implement, Release, and Maintain API Products That Meet User Needs*, Birmingham, 2024.
- REHR (Jean-Paul) et BEAULIEU (Marie-Anne Polo de), « Thesaurus Exemplorum Medii Aevi : une base de données collaborative sur les exempla médiévaux », *Humanités numériques*, 4-2 (2021), DOI : 10.4000/revuehn.2630.
- THIERRY (Benjamin), « Jean-Philippe Genet et Andrea Zorzi (dir.), Les Historiens et l'informatique : un métier à réinventer », *Histoire, Économie et Société*, 33-3 (2014), p. 119-120, JSTOR : 24467721, URL : <https://www.jstor.org/stable/24467721> (visité le 05/04/2024).
- TUBACH (Frederic C.), *Index exemplorum : a handbook of medieval religious tales*, Helsinki, Finlande, 1969.
- WESSELS (Bridgette), WADHWA (Kush) et FINN (Rachel L.), *Open Data and the Knowledge Society*, Amsterdam, 2017, DOI : 10.1515/9789048529360.

Webographie

- 17 *Linking, Segmentation, and Alignment - The TEI Guidelines*, URL : <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html> (visité le 15/08/2024).
- 20 *Graphs, Networks, and Trees - The TEI Guidelines*, URL : <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/GD.html> (visité le 31/08/2024).
- Bases de données textuelles du CRH*, URL : <http://crh.ehess.fr/index.php?7473> (visité le 26/08/2024).
- Cytoscape : An Open Source Platform for Complex Network Analysis and Visualization*, URL : <https://cytoscape.org/> (visité le 31/08/2024).
- Gephi - The Open Graph Viz Platform*, URL : <https://gephi.org/> (visité le 31/08/2024).
- Informations sur ThEMA*, URL : <https://thema.huma-num.fr/about/> (visité le 26/08/2024).
- Labex Hastec (EPHE) / Un laboratoire d'excellence de PSL - Paris Sciences et Lettres, une université internationale, au cœur de Paris*, URL : <https://labexhastec.ephe.psl.eu/> (visité le 26/08/2024).
- Le Pôle numérique*, URL : <https://www.irht.cnrs.fr/fr/qui-sommes-nous/1-organisation-de-1-irht/pole-numerique> (visité le 26/07/2024).
- Les récits exemplaires de la moniale chaste et bavarde dans ThEMA*, URL : <https://thema.huma-num.fr/search?term=&collection=all&author=all&context=all&keyword=&keyword=&keyword=&tubach=TUB723&tubach=&tubach=&allegory=all> (visité le 01/09/2024).
- Page dédiée au mot exemplum dans Wikitionnaire*, juin 2024, URL : <https://fr.wiktionary.org/w/index.php?title=exemplum&oldid=35038953> (visité le 02/09/2024).
- Partenaires du projet Biblissima*, URL : <https://projet.biblissima.fr/en/community/founding-partners> (visité le 02/09/2024).
- Portail Biblissima*, URL : <https://www.irht.cnrs.fr/fr/ressources/sites-web-outils-corpus/biblissima> (visité le 29/07/2024).
- Récit exemplaire utilisé pour la comparaison de textes*, URL : http://sourcencyme.irht.cnrs.fr/encyclopedie/summa_de_exemplis_ac_similitudinibus_rerum_ed_anvers_1609?citid=cit_idp119842416 (visité le 01/09/2024).
- SourcEncyMe*, URL : <http://sourcencyme.irht.cnrs.fr/> (visité le 29/07/2024).

SourceEncyMe Sources des Encyclopédies Médiévales, corpus annoté, URL : <http://sourcencyme.irht.cnrs.fr/historique> (visité le 26/08/2024).

Sources et données de la recherche, URL : <http://crh.ehess.fr/index.php?6813> (visité le 26/08/2024).

ThEMA, URL : <https://thema.huma-num.fr/> (visité le 29/07/2024).

Introduction

En 2011, lors d'un colloque sur les interactions entre le numérique et l'histoire, les chercheurs ont constaté que « la discipline historique vit actuellement au moyen âge de son rapport à l'ordinateur, à la fois définitivement coupée des temps anciens (quand on travaillait « sans »), mais également marquée par le gothique des initiatives, désordonnées et tâtonnantes ».¹.

Depuis, la situation a évolué vers une uniformisation et un regroupement des productions numériques appliquées aux sciences sociales. Cela se manifeste, par exemple, par la création du portail Biblissima, qui propose un accès unifié à un ensemble de données numériques sur les manuscrits, incunables et imprimés anciens, provenant des neuf équipes partenaires du consortium Biblissima². De plus, l'importance croissante de l'interopérabilité et du « Linked Open Data » dans le champ des humanités numériques, telle que constatée par Terhi Nurmikko-Fuller³, témoigne également de cette évolution.

Cependant, il reste encore du travail à accomplir concernant les anciennes bases de données, qui n'ont pas été conçues à l'origine pour être interconnectées avec d'autres. C'est dans cette perspective que s'inscrit mon stage. Grâce au financement du LabEx HaStec, mon objectif a été de mettre en relation deux bases de données : ThEMA⁴ et SourcEncyMe⁵.

I. Présentation des bases de données ThEMA et SourcEncyMe

ThEMA peut se définir comme une base de données multimédia et multilingue sur des anecdotes exemplaires provenant principalement de l'Occident médiéval. Elle tire son origine d'un groupe de travail qui progressivement a enregistré des données au format Mi-

1. Benjamin Thierry, « Jean-Philippe Genet et Andrea Zorzi (dir.), Les Historiens et l'informatique : un métier à réinventer », *Histoire, Économie et Société*, 33–3 (2014), p. 119–120, JSTOR : 24467721, URL : <https://www.jstor.org/stable/24467721> (visité le 05/04/2024).

2. Portail Biblissima, URL : <https://www.irht.cnrs.fr/fr/ressources/sites-web-outils-corpus/biblissima> (visité le 29/07/2024).

3. Terhi Nurmikko-Fuller, *Linked Data for Digital Humanities*, London, 2023.

4. ThEMA, URL : <https://thema.huma-num.fr/> (visité le 29/07/2024).

5. SourcEncyMe, URL : <http://sourcencyme.irht.cnrs.fr/> (visité le 29/07/2024).

crosoft Word sur les récits exemplaires qui les intéressaient⁶. Ces fichiers sont repris dans les années 2000 par Marjorie Burghart pour créer une base de données PHP⁷ MySQL⁸ avec une interface de saisie sous Microsoft Access qui permettait, hors ligne, le travail des indexateurs⁹. Cette interface est ensuite remplacée par un site Web en 2005. Il est directement lié à une base de données MySQL. Ce type de bases permet une collaboration dynamique dans le contrôle de la qualité et la collecte des données. Le lancement de ce site a été rendu possible grâce aux efforts de Marjorie Burghart et de Pascal Collomb (EHESS), qui s'est intégré plus tardivement à l'équipe¹⁰. En 2015, Marjorie Burghart a suggéré de transférer le projet ThEMA à Huma-Num, une très grande infrastructure de recherche (TGIR) pour les sciences humaines financée par le CNRS, dans le but de garantir la sécurité à long terme de l'environnement du serveur¹¹. Par la suite, la décision a été prise de développer un schéma XML-TEI pour les données de ThEMA et de les intégrer à un serveur d'application eXist-db, supervisés par Jean-Paul Rehr (université Lumière Lyon 2 et CIHAM), un outil qui a été lancé à l'automne 2019¹².

ThEMA s'inscrit dans une longue tradition de bases de données créées par l'EHESS et le CRH. Depuis les années 1960, le CRH consacre une part importante de son activité à la constitution d'outils de recherche, mis à disposition de l'ensemble de la communauté savante¹³. ThEMA est même la première base de données élaborée au sein de l'EHESS à avoir été mise en accès direct et libre¹⁴. Actuellement, cette base est gérée d'un point de vue technique par Jean-Paul Rehr et Jean-Damien Généro, et d'un point de vue scientifique par Marie-Anne Polo de Beaulieu. Elle est intégrée au collectif « Sources et données de la recherche », créé en 2019 pour accompagner les projets numériques du CRH¹⁵.

De son côté, SourcEncyMe se définit comme un corpus des encyclopédies médiévales latines, où sont progressivement identifiées les sources grecques, arabes et latines de la pensée scientifique et philosophique. Cette base trouve ses origines dans l'« Atelier Vincent

6. Jean-Paul Rehr et Marie-Anne Polo de Beaulieu, « Thesaurus Exemplorum Medii Aevi : une base de données collaborative sur les exempla médiévaux », *Humanités numériques*, 4–2 (2021), DOI : 10.4000/revuehn.2630.

7. Il s'agit d'un langage de script côté serveur utilisé pour le développement web. PHP peut être intégré dans le code HTML et est souvent utilisé pour créer des pages web dynamiques. Il permet de se connecter à une base de données MySQL pour manipuler les données (ajouter, modifier, supprimer, lire).

8. Il s'agit d'un système de gestion de bases de données relationnelles (SGBDR) très populaire. Il permet de stocker, organiser et récupérer des données de manière efficace. MySQL utilise le langage SQL (Structured Query Language) pour gérer les bases de données.

9. Marjorie Burghart, « L'informatisation du ThEMA », *Les Cahiers du Centre de Recherches Historiques*, 35 (2005), DOI : 10.4000/ccrh.3025.

10. J.P. Rehr et M.A. P. d. Beaulieu, « Thesaurus Exemplorum Medii Aevi... ».

11. *Ibid.*

12. *Ibid.*

13. *Bases de données textuelles du CRH*, URL : <http://crh.ehess.fr/index.php?7473> (visité le 26/08/2024).

14. *Informations sur ThEMA*, URL : <https://thema.huma-num.fr/about/> (visité le 26/08/2024).

15. *Sources et données de la recherche*, URL : <http://crh.ehess.fr/index.php?6813> (visité le 26/08/2024).

de Beauvais », qui, dès les années 1980, sous l’initiative du Doyen Jean Schneider et avec les efforts de Monique Paulmier-Foucart, assistée de Marie-Christine Duchenne, a entrepris la mise en ligne du *Speculum historiale* de Vincent de Beauvais, sur le serveur du laboratoire ATILF à Nancy¹⁶. Le projet a pris de l’ampleur à partir de 2007 avec le développement d’un corpus de toutes les encyclopédies latines médiévales, se transformant en 2010 en une base de données et une plateforme collaborative permettant l’enrichissement des données scientifiques¹⁷. Le site SourcEncyMe.irht.cnrs.fr a été ouvert au public le 24 février 2016 à l’Institut de recherche et d’histoire des textes (IRHT, CNRS)¹⁸. À partir de 2020, SourcEncyMe a subi une transformation technique, passant à un corpus intégré en XML-TEI avec une base de données unifiée en XML, associée à la création d’une interface de balisage¹⁹.

Aujourd’hui, SourcEncyMe est l’une des nombreuses bases de données gérées par l’IRHT, un institut qui a été parmi les premiers à adopter les progrès techniques du numérique pour le traitement des sources historiques²⁰. La base fait partie du pôle numérique de l’institut, lequel vise à assurer l’interopérabilité, la diffusion, la valorisation et l’enrichissement des données issues de la recherche²¹. D’un point de vue technique, SourcEncyMe est gérée par Henri Seng et Emmanuelle Kuhry, tandis que la gestion scientifique est assurée par Isabelle Draelants.

II. Motivations et enjeux de l’interconnexion des bases de données

Bien que ThEMA et SourcEncyMe soient des entités distinctes avec des histoires, des institutions d’attache et des sources différentes, plusieurs points communs permettent d’envisager leur interconnexion.

D’un point de vue historique, les recueils d’*exempla* et les encyclopédies participent tous deux à l’encyclopédisme du XIII^e siècle²². Par exemple, un même auteur, comme Thomas de Cantimpré, a pu rédiger à la fois une encyclopédie (*De natura rerum*) et

16. *SourcEncyMe Sources des Encyclopédies Médiévales, corpus annoté*, URL : <http://sourcencyme.irht.cnrs.fr/historique> (visité le 26/08/2024).

17. Emmanuelle Kuhry, *Quels traitements numériques pour les corpus encyclopédiques ? Les projets SourcEncyMe et Mythologia en dialogue (retour sur la séance du séminaire « Les p’tits déj’ ‘Humanités numériques’ de l’IRHT »)*, 14/01/22, DOI : 10.58079/nmg9.

18. Isabelle Draelants et E. Kuhry, « Encyclopédies médiévales en milieu numérique : Les nouveaux enjeux de SourcEncyMe pour le traitement des auctoritates », dans *Dix ans de corpus d'auteurs*, dir. Fatima Idmhand et Ioana Marasescu-Galleron, éditions des archives contemporaines, Paris, 2022, p. 155-178, DOI : 10.17184/eac.5479.

19. *Ibid.*

20. Louis Holtz, *L’IRHT, au fil des ans*, 2015, DOI : 10.58079/qhxe.

21. *Le Pôle numérique*, URL : <https://www.irht.cnrs.fr/fr/qui-sommes-nous/1-organisation-de-l-irht/pole-numerique> (visité le 26/07/2024).

22. Jacques Le Goff, « Pourquoi le XIII^e siècle a-t-il été plus particulièrement un siècle d’encyclopédisme ? », dans *L’enciclopedia medievale*, dir. Michelangelo Picone, Ravenna, 1994, p. 23-40.

un recueil de récits exemplaires (*Bonum universale de apibus*). Ces œuvres partagent une rhétorique commune visant à décoder l’œuvre de Dieu sur Terre. De plus, les encyclopédies et les recueils d’*exempla* se réfèrent respectivement à leur contenu. Par exemple, des éléments du *Speculum majus* de Vincent de Beauvais ont été repris dans 28 recueils de récits exemplaires entre le XIII^e et le XVe siècle²³.

D’un point de vue technique, les textes des deux bases sont encodés en XML-TEI, un format de balisage spécifique pour les documents textuels. Ce format permet à l’information d’avoir exactement le même aspect pour tous les logiciels qui l’utilisent, même si la structure du XML peut différer²⁴. De plus, les systèmes de gestion de bases de données (SGBD) utilisés, BaseX et eXist-db, sont très similaires. Ces SGBD stockent les données XML sous forme native, utilisent XQuery comme langage de requête principal, et sont dotés de licences offrant une grande liberté d’utilisation et de modification. Enfin, même l’interface visuelle des deux bases présente des ressemblances :



FIGURE 1 – Page d’accueil de SourcEncyMe

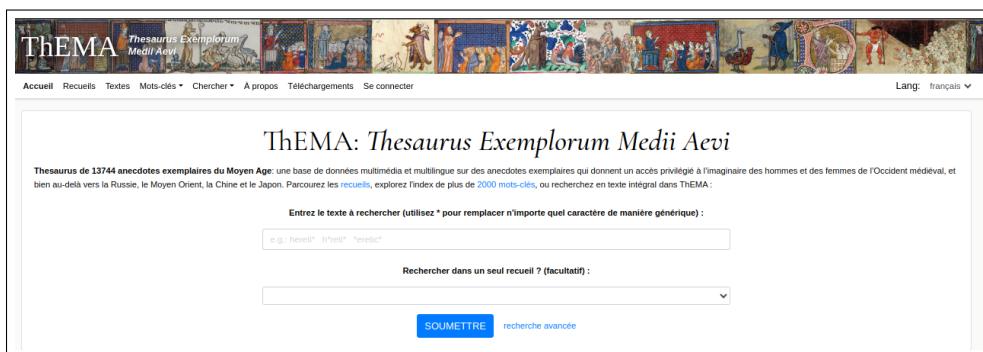


FIGURE 2 – Page d’accueil de ThEMA

D’un point de vue scientifique, les deux projets partagent une vocation commune :

23. Jacques Berlioz et Marie-Anne Polo de Beaulieu, « Les recueils d’exempla et la diffusion de l’encyclopedisme médiéval », dans *L’enciclopedismo medievale*, dir. Michelangelo Picone, Ravenna, 1994, p. 179-212.

24. Lou Burnard, *Qu'est-ce que la Text Encoding Initiative ?*, Marseille, 2015, URL : <https://books.openedition.org/oep/1237> (visité le 11/04/2024).

rendre accessible un patrimoine ancien et précieux pour comprendre le répertoire des connaissances médiévales et les informations sur la vie matérielle au Moyen Âge. Ils intègrent les textes dans une histoire de la transmission des savoirs, en accordant une attention particulière aux métadonnées qui indiquent la source des textes et les textes apparentés.

D'un point de vue institutionnel, l'EHESS et l'IRHT entretiennent de bonnes relations²⁵. Des chercheurs des deux institutions participent à des projets communs, bien qu'ils soient parfois en concurrence. C'est le cas de Marie-Anne Polo de Beaulieu pour ThEMA et d'Isabelle Draelants pour SourcEncyMe, qui sont toutes deux impliquées dans le LabEx HaStec, un laboratoire d'excellence se concentrant sur les savoirs, les techniques et les croyances, des thèmes bien représentés dans les deux bases de données²⁶. Ce type de projet est crucial car il favorise la création de liens entre les institutions. Grâce au financement du LabEx HaStec, j'ai ainsi pu réaliser ce stage, permettant ainsi aux équipes des deux projets de se rencontrer. Le LabEx HaStec a également financé une post-doctorante, Élisa Lonati, qui travaillera aussi à cheval entre ThEMA et SourcEncyMe, assurant la continuité des liens établis entre les deux équipes.

III. Présentation de la problématique et du plan

L'interopérabilité est devenue un enjeu majeur en sciences sociales ces dernières années, comme mentionné précédemment. Elle est essentielle pour rendre les données de recherche accessibles dans un environnement universitaire fragmenté, où les intérêts scientifiques varient, les sources historiques sont dispersées, et les bases de données souvent isolées. Toutefois, grâce à la convergence des efforts de certains chercheurs, cet objectif reste réalisable. La question que je vais explorer est donc la suivante : comment connecter efficacement deux systèmes de gestion de bases de données (SGBD) en XML-TEI, qui, bien que distincts, présentent des similarités ?

Pour répondre à cette problématique, j'examinerai d'abord les différentes options disponibles et la modalité choisie pour mettre en relation les deux bases de données. Ensuite, je détaillerai la mise en œuvre concrète de cette connexion. Enfin, je prendrai du recul sur ce travail de connexion pour évaluer ses apports et ses perspectives.

25. L. Holtz, « L'institut de recherche et d'histoire des textes (IRHT) », *Les Cahiers du Centre de Recherches Historiques*, 36 (2005), DOI : 10.4000/ccrh.3046.

26. *Labex Hastec (EPHE) / Un laboratoire d'excellence de PSL - Paris Sciences et Lettres, une université internationale, au cœur de Paris*, URL : <https://labexhastec.ephe.psl.eu/> (visité le 26/08/2024).

Chapitre 1

Réflexions sur les différentes possibilités d’interconnexion

Pour établir une interconnexion efficace entre les bases de données ThEMA et SourcEncyMe, trois solutions ont été envisagées : la fusion des bases de données, la création d’une API, et l’ajout de liens simples dans les fichiers XML-TEI. Chacune de ces solutions présente des avantages et des inconvénients, tant sur le plan technique qu’institutionnel et scientifique.

I. Solution radicale : fusionner les deux bases de données

L’une des solutions les plus directes pour relier ThEMA et SourcEncyMe consiste à fusionner les deux bases de données en un projet unique. Cette option pourrait être envisagée en regroupant les deux bases sous l’égide d’une infrastructure commune, telle que Biblissima, puisque l’IRHT et l’EHESS sont déjà partenaires de ce projet¹. Une autre possibilité serait de les fusionner directement dans ThEMA ou SourcEncyMe.

La fusion des deux bases de données présente plusieurs avantages. Tout d’abord, elle élimine la nécessité de développer une API, ce qui pourrait simplifier considérablement le processus d’interconnexion. En outre, en consolidant les deux systèmes de gestion de bases de données (SGBD), la maintenance serait également simplifiée, car une seule base de données nécessiterait une gestion continue.

Cependant, cette solution présente des inconvénients. Sur le plan institutionnel, la fusion impliquerait que l’une des deux institutions, l’EHESS ou l’IRHT, prenne en charge l’ensemble du travail de gestion des deux bases, ce qui pourrait poser des problèmes de responsabilité et de gouvernance. D’un point de vue historique, la fusion risquerait de

1. Partenaires du projet Biblissima, URL : <https://projet.biblissima.fr/en/community/founding-partners> (visité le 02/09/2024).

brouiller la distinction entre les deux types de sources que sont les encyclopédies et les récits exemplaires. Enfin, sur le plan scientifique, bien que les deux bases poursuivent des objectifs similaires en termes de valorisation du patrimoine médiéval, ThEMA se concentre sur la diversité des récits exemplaires², tandis que SourcEncyMe se focalise sur la transmission des textes encyclopédiques³. Ces différences essentielles seraient potentiellement compromises par une fusion.

Face à ces défis, il a été nécessaire de considérer d’autres solutions moins radicales et plus respectueuses des spécificités institutionnelles et scientifiques de chaque base.

II. Solution idéale : développer une API

La création d’une API (Interface de Programmation d’Applications) permettrait de connecter les deux bases tout en les maintenant distinctes. Une API est un ensemble de spécifications techniques permettant à des logiciels de communiquer entre eux en définissant des protocoles, méthodes, et formats de données. Cette solution offrirait un point d’accès centralisé pour interroger les deux bases, et permettrait de manipuler des données provenant de schémas XML différents, même si ces schémas évoluent⁴.

L’avantage de cette approche est de pouvoir, par exemple, rechercher un *exemplum* dans SourcEncyMe et récupérer les récits exemplaires similaires dans ThEMA en utilisant l’indexation de l’*Index exemplorum* de Frederic Tubach⁵, sans avoir à recréer un système d’indexation des *exempla* dans SourcEncyMe⁶.

Cette solution maintient la distinction historique entre les sources et respecte les spécificités scientifiques et institutionnelles des deux bases. Cependant, elle a été écartée en raison de la complexité technique et du temps limité du stage. ThEMA dispose déjà d’une API, mais SourcEncyMe n’en avait pas, et je n’avais pas accès direct au code source, seulement aux fichiers contenant les encyclopédies.

Ces contraintes nous ont conduits à opter pour une dernière option : l’ajout de simples liens dans les fichiers XML.

III. Un compromis : l’ajout de liens simples

La solution finalement retenue pour l’interconnexion des deux bases de données a consisté à ajouter de simples liens dans les fichiers XML-TEI. Cette approche, bien que

2. ThEMA...

3. SourcEncyMe...

4. Bruno Pedro, *Building an API Product : Design, Implement, Release, and Maintain API Products That Meet User Needs*, Birmingham, 2024.

5. Frederic C. Tubach, *Index exemplorum : a handbook of medieval religious tales*, Helsinki, Finlande, 1969.

6. L’ouvrage de Tubach, publié en 1969, propose une classification et un catalogue des récits exemplaires en fonction de leurs thèmes, motifs et caractéristiques. Il est intégré à ThEMA.

techniquement moins ambitieuse que la création d’une API ou la fusion des bases, présente plusieurs avantages.

L’ajout de liens est une tâche relativement simple à mettre en œuvre. Il s’agit d’insérer des balises XML-TEI qui pointent vers des données similaires présentes dans les deux bases. Ces liens peuvent ensuite être activés dans le HTML des bases de données, permettant aux utilisateurs de naviguer facilement entre ThEMA et SourcEncyMe.

Cette solution présente l’avantage d’être facilement réalisable dans le cadre d’un stage court et de ne nécessiter qu’une formation technique limitée. De plus, elle évite les conflits institutionnels, scientifiques et historiques, car les deux bases restent distinctes tout en étant reliées. Ainsi, cette méthode offre un bon compromis, permettant d’interconnecter les deux bases de manière efficace sans compromettre leur indépendance.

Chapitre 2

Stratégies pour l'intégration des liens dans les deux bases de données

Avant d'établir les liens entre les bases de données, il est essentiel de préparer correctement ces bases. Cela inclut la vérification de la qualité et de la pérennité des données, ainsi que la planification précise de l'insertion des liens dans les fichiers XML et des données ou métadonnées à associer.

I. Vérification de la qualité des données et de leur pérennité

Je me suis principalement chargé de la vérification de la qualité des données et de la pérennité de ThEMA, car je n'ai pas eu de tâches similaires du côté de SourcEncyMe.

J'ai d'abord rédigé une documentation détaillée sur le fonctionnement et l'installation locale de la base de données. Ce travail, présenté dans les annexes A et B, est crucial pour éviter l'abandon des bases de données lorsque les chercheurs responsables prennent leur retraite. La disparition éventuelle de la base compromettrait la valeur des liens créés. Ce point est d'autant plus pertinent avec le départ prochain à la retraite de Marie-Anne Polo de Beaulieu, directrice scientifique du projet. Il est impératif que cette documentation facilite le transfert de connaissances aux futurs chercheurs ou ingénieurs d'études qui reprendront ThEMA. Actuellement, Jean-Paul Rehr, le créateur de la base, en assure seul la gestion. Il est donc essentiel que ses connaissances soient préservées.

J'ai également résolu un problème lié à l'affichage des *exempla*. Initialement, les *exempla* n'étaient pas classés selon l'ordre du recueil, mais étaient triés en fonction du numéro du fichier XML. Par exemple, un *exemplum* numéroté 10 pourrait apparaître avant un *exemplum* 5.2 si ce dernier avait été ajouté ultérieurement, car son fichier XML avait été généré plus tard. De plus, des incohérences liées aux différentes méthodes d'indexa-

tion utilisées par les chercheurs ont été identifiées : certains *exempla* étaient classés avec des notations variées telles que 1.b ; 1b ; 1, 5, 6. Cette diversité de formats empêchait un classement croissant uniforme. La réorganisation était donc nécessaire pour garantir que les utilisateurs puissent retrouver facilement les *exempla* recherchés. Sans cette réorganisation, les liens vers SourcEncyMe risquaient de devenir moins accessibles.

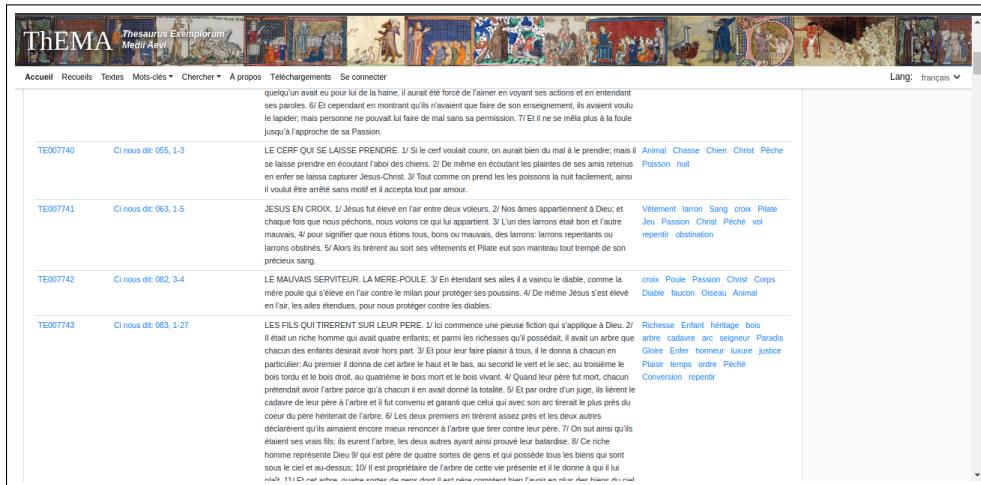


FIGURE 2.1 – Récits exemplaires mal classés

Pour résoudre ce problème, j'ai d'abord développé un code en Python (annexe C). Ce code extrait les numéros des récits exemplaires et convertit les lettres présentes dans la numérotation en chiffres. Cependant, cette méthode a présenté des inconvénients : elle nécessitait de modifier de nombreux types de numérotations et de réintégrer les fichiers XML modifiés dans la base, alors que seulement 30 % des récits étaient mal classés.

Une approche plus simple a donc été développée pour identifier uniquement les recueils présentant des erreurs de classement et les corriger sans tenter d'uniformiser les numéros des récits. De plus, ce second code a été écrit en XQuery (annexe D) afin de s'intégrer directement dans la base de données. Ainsi, en cas de nouvelles erreurs d'indexation, elles seront automatiquement traitées, éliminant la nécessité de relancer le code Python et de réintégrer les données dans la base.

II. Identifications des liens entre les bases de données

J'ai identifié des similitudes entre les deux bases de données, tant au niveau des données que des métadonnées. Pour les données, c'est-à-dire les éléments bruts susceptibles de fournir des informations (ici, tu texte), les types de connexions possibles sont les suivants :

- Quand un *exemplum* d'un recueil de récits exemplaires utilise un morceau d'une encyclopédie.

- Quand un *exemplum* est repéré dans une encyclopédie et qu'un recueil d'*exempla* contient un *exemplum* similaire ou si on le retrouve dans la classification de l'*Index exemplorum* de Frederic Tubach.
- Quand un *exemplum* est identifié dans une encyclopédie et qu'il est indexé dans ThEMA.

Pour ce qui est des métadonnées, qui fournissent des informations sur d'autres données, voici les éléments pouvant être connectés :

- Les mementos (fiches de présentation des auteurs ou des œuvres) des deux bases de données et leur contenu respectif peuvent être connectés :

The screenshot shows the 'Recueil' section of the ThEMA interface. It displays a list of entries under the heading 'Liste des exempla (9)'. The first entry is 'Astotus Viennensis'. The interface includes various filters and search fields, such as 'Auteur', 'Langue', 'Editions Imprimées', and 'Indexation par'. At the bottom, there is a list of terms and a note about the last update.

FIGURE 2.2 – Memento dans ThEMA

The screenshot shows the 'Memento' section of the SourcEncyMe interface for the author 'Vincentius Bellovacensis'. It provides biographical information, including his name, date of death (1264), and occupation (Dominican). It also details his work on the 'Speculum maius', his relationship with King Louis IX, and his role in the reform of the Cistercians. The interface includes tabs for 'Parcours', 'Fonds documentaire', and 'Bibliographie'.

FIGURE 2.3 – Memento dans SourcEncyMe

- Les sources des récits exemplaires mentionnées lors de l'indexation sur ThEMA peuvent être connectées aux mementos, comme illustré avec Cicéron :

Exemplum

Collection: Liber de introductione loquendi [Vecchio, 1998] (TC0137)

Auteur: Philippe de Ferrare

Exemplum context: L. I, cap. 144.

Summary: (Fr.) La vie des riches est dangereuse. Un diable fait s'asseoir Denis, qui avait obtenu des grands, honneurs et richesses, sur un trône en or et lui fait préparer un dîner plein de tous les délices possibles. Il lui place également une épée liée à un crin de cheval sur la tête. Denis ne bougea pas, ayant peur de l'épée. Le diable expliqua alors à Denis que telle était la vie du diable, vie que Denis avait crue meilleure. (It.) La vita dei ricchi è pericolosa. Un diavolo fece sedere Dioniso, che era arrivato ai maggiori onori e ricchezze, su un trono aureo e gli fece preparare un pranzo pieno di tutte le delizie possibili, ma gli fece anche porre una spada legata per un crine di cavallo sopra la sua testa. Dioniso allora non si mosse per paura della spada. Allora il diavolo spiegò a Dioniso che tale posizione era la vita del diavolo che Dioniso aveva reputato ottima.

Sources: Cicero, Tusculanae disputationes, 5, 21, 61-62.

Keywords: sword meal Devil rich education

FIGURE 2.4 – Récit exemplaire utilisant un texte de Cicéron

Memento

Nom actualisé de l'auteur : Cicero Marcus Tullius

Oeuvres authentiques :

- Cato Maior de senectute
- De divinatione Ciceronis
- De domo sua ad pontifices oratio
- De fato Ciceronis
- De finibus bonorum et malorum Ciceronis
- De haruspicium responso oratio
- De legibus Ciceronis
- De officiis Ciceronis
- De oratore Ciceronis
- De provinciis consularibus oratio
- De re publica Ciceronis
- De universo Ciceronis
- Hortensius Ciceronis
- In L. Serio Catilinam orationes Ciceronis

FIGURE 2.5 – Memento de Cicéron

- Les noms de lieux, de personnes et les thématiques présents dans les mots-clés de ThEMA peuvent être reliés aux textes de SourcEncyMe, comme le montre l'exemple avec Jérusalem :

Exemplum

Recueil: Ci nous dit ou Composition de l'Ecriture sainte [Blangez, 1979-88] (TC0131)

Auteur: anon.

Résumé: (Fr.) JUDITH TUE HOLOPHERNE. 1/ La belle Judith envira Holopherne et le tua pendant son sommeil parce qu'il voulait détruire Jérusalem; et elle suspendit sa tête sur la muraille de la cité. 2/ Quand ses soldats virent cela, ils se débandèrent. 3/ C'est ainsi que nous fûmes délivrés d'enfer par la vierge Marie que figurait Judith; 4/ Les diables perdirent leur pouvoir par elle et par son fils; et nous fûmes délivrés d'enfer tout comme ceux de Jérusalem furent délivrés d'Holopherne par la belle Judith.

Sources: Jud. 13-15

Commentaire: Exemplum biblique.

Mots-clés: Judith Holopherne Meurtre Jérusalem décapitation

FIGURE 2.6 – Mention de Jérusalem dans ThEMA

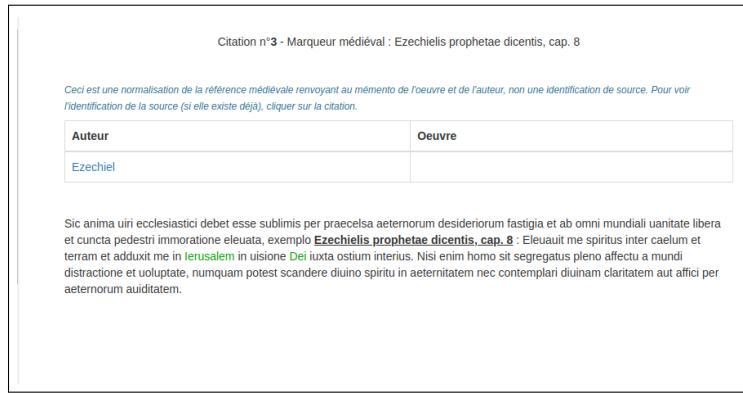


FIGURE 2.7 – Mention de Jérusalem dans SourcEncyMe

Parmi les options de lien disponibles, j'ai décidé de me concentrer principalement sur les données. Cette décision repose sur plusieurs considérations : c'était un des objectifs principaux définis dans le livret de stage, et les utilisateurs s'intéressent principalement aux données, cherchant avant tout à accéder aux résumés de ThEMA et aux textes édités dans SourcEncyMe. Les mementos et les mots-clés, bien que facilitant les recherches thématiques, sont secondaires par rapport aux données essentielles pour la recherche historique. De plus, une encyclopédie testée avait déjà fait l'objet d'un repérage préalable d'*exempla* par une autre chercheuse de l'IRHT, Mara Calloni, et l'indexation des encyclopédies dans ThEMA a facilité ce travail.

En ce qui concerne les mots-clés, les noms de lieux et de personnes, le travail aurait été plus complexe, car SourcEncyMe n'a effectué qu'un repérage partiel à ce niveau dans les textes, contrairement à ThEMA. Quant aux liens vers les mementos, ils n'étaient pas prioritaires par rapport aux liens entre les textes, même s'ils auraient pu être établis assez facilement.

III. Intégration des liens XML dans les bases de données

Les textes des deux bases de données sont stockés dans des fichiers XML, ce qui a nécessité une réflexion sur l'insertion des liens et la recherche d'informations permettant de créer des liens. La gestion de ces liens a varié en fonction des spécificités de chaque base.

1. Recherche des emplacements et des informations dans les XML pour la création des liens dans ThEMA

Pour ThEMA, l'intégration des liens a été relativement simple. Les fichiers XML, un par *exemplum*, comportaient déjà des balises spécifiques pour les liens, comme illustré

dans l'exemple suivant :

```

1   <sourceDesc>
2       <list type="source_details">
3           <item type="source_text">
4               <p>
5               </p>
6           <item type="sources">
7               <p>
8               </p>
9           <item type="exemplum_context">
10              <p>Feria quarta primae hebdomadae. Sermo I.</p>
11          </item>
12          <item type="commentary">
13              <p>
14              </p>
15          <item type="allegory">y</item>
16      </list>
17      <listBibl>
18          <bibl type="manuscripts-editions" corresp="Z-IU7WR86C" n="t. I, p. 66-67"/>
19      </listBibl>
20      <list type="links">
21          <item type="link" corresp="http://sermones.net/thesaurus/document.php?id=jvor_210">Sermones.net</item>
22      </list>
23      <list type="linked_exempla"/>
24  </sourceDesc>

```

Jacobus de Voragine, Sermones aurei [ed. Clutius, 1760]: p. 27b

État: publié
Identifiant: TE002904
Comment citer: « TE002904 (Sermones aurei · Quadragesimale: p. 27b) », *Thesaurus exemplorum medi aevi*, <https://thema.huma-num.fr/exempla/TE002904> (consulté 2024-08-14).

Exemplum

Recueil: Sermones aurei - Quadragesimale (TC0011)
Auteur: Jacobus de Voragine
Contexte de l'exemplum: Feria quarta primae hebdomadae. Sermo I.

Résumé: (Fr.) L'odeur de la bonne réputation du Christ est assimilée aux odeurs du cédré, de la myrrhe et de la vigne qui font fuir les crapauds (représentation de la luxure), les serpents (représentation de l'avarice) et les vers (représentation de l'orgueil). Les crapauds naissent souvent des organes génitaux des cadavres des hommes.

Mots-clés: Christ Avarice Crapaud luxure Serpent Orgueil Odeur Ver vigne myrrhe cédre

Références

Manuscrits / éditions: JACOBUS DE VORAGINE, 1764. *Sermones aurei*, Figarol, . t. I, p. 66-67.
Liens: Sermones.net

FIGURE 2.8 – Localisation du lien dans ThEMA

Les références aux encyclopédies dans ThEMA se trouvent dans les champs « sources » et « textes apparentés ». Pour les récupérer, il faut naviguer dans les balises suivantes : <sourceDesc>, puis <list>, puis <item> avec l'attribut type="source", et enfin <p>. De même, pour les textes apparentés, il faut se rendre dans <sourceDesc>, puis <listBibl>, et enfin <bibl type="related_texts"/>.

```

1   <sourceDesc>
2       <list type="source_details">
3           <item type="source_text">
4               <p>
5               </p>
6           <item type="sources">
7               <p>Vincent de Beauvais, Speculum historiale, 13.50, dans Speculum quadruplex, vol. 4, 522a.</p>
8           </item>
9       <list>
10      <sourceDesc>

```

```

1 </sourceDesc>
2     <listBibl>
3         <bibl type="related_texts" corresp="Z-E46747N6" n="p. 584"/>
4         <bibl type="tubach" corresp="Z-WLZ7CBVC" n="TUB951"/>
5     </listBibl>
6     <list type="links"/>
7     <list type="linked_exempla"/>
8 </sourceDesc>

```

The screenshot shows a detailed view of a source entry in the ThEMA interface. At the top, there is a header with fields like 'État' (published), 'Identifiant' (TE017648), and 'Comment citer' (with a link to the Thesaurus exemplorum mediæ aevi). Below this is a section titled 'Exemplum' containing various details such as 'Recueil' (De oculo morali (TC0159)), 'Auteur' (Petrus Lemovicensis), and 'Contexte de l'exemplum' (Chapter 12, 1A, Prima instructio prelatorum. Quod prelati sicut dignitate sic precellere debent sanctitate). There is also a 'Résumé' (summary in French) and a 'Sources' section listing Vincent de Beauvais, Speculum historiale, 13.50, dans Speculum quadruplex, vol. 4, 522a. A 'Commentaire' section provides biographical information about Hélène or sainte Hélène. The 'Mots-clés' section includes terms like Mère, Constantin, empereur, Fils, injustice, oeil, regard, and Jugement. At the bottom, there is a 'Références' section with a note about a translation by Peter of Limoges.

FIGURE 2.9 – Champ source dans ThEMA

2. Placement des liens et des visualisation des *exempla* dans les XML de SourcEncyMe

En revanche, dans SourcEncyMe, il n'existe pas de balises spécifiques pour intégrer les liens vers ThEMA. Chaque encyclopédie est intégrée dans un fichier XML unique, comme montré dans l'annexe E. Initialement, j'avais envisagé d'ajouter de nouvelles balises `<ref>` avec un attribut spécifique pour les liens, en conformité avec les TEI Guidelines¹. Ce lien aurait été intégré dans la balise `<cit>` du XML, car dans SourcEncyMe, chaque citation d'un auteur est placée dans une balise `<cit>`. Voici la structure envisagée :

```

1 <cit xml:id="cit_idp103803440">
2     <bibl>
3         <author ref="#gregorius_nazianzenus">Gregorius Nazianzenus</author>
4     </bibl>
5     <quote>Et hinc est quod venerabilis doctor<seg type="marqueur">Gregorius Nazianzenus</seg>cuius (testante<seg type="marqueur">Hieronymus</seg>)
       tanta auctoritatis fuit; ut nullus unquam eius dictis calumniam inferre presumpserit; Immo insuper (ut testatur<seg type="marqueur">Rufinus</seg>) tanta fuit eius auctoritas apud ecclesias Christi, ut esse putaretur hereticus qui illi fuisse ausus in aliquo contraire. Hic itaque doctor hanc sibi consuetudinem fecerat, ut quicquid eius occurreret oculis de rebus exterioribus, interioribus anime moribus adaptaret; Sic enim referat ipse de se in suo apologetico, dicens; Mos mihi est ad meipsum singula queque recovare, precipue si quies mihi, sit, et silentium si vacet, animi moribus adaptare quod videtur in oculis. Unde et oculorum visio est mihi mentis eruditio. Huius igitur exemplum laudabile sequi oportet, si abundare volunt copia exemplorum. Que copia illis si affuerit, precipue de rebus extrinsecis que nobis sunt in aperto, et etiam de mirandis operibus que continue natura producit, vel adinvenit humana industria; non eos tantum apud curiosos auditores faciet gratiosos, quos mira nature opera vel humane inventionis studia narrata et patefacta delectant, sed etiam apud vulgus et simplices fructuosos et acceptos constituet, dum per exempla ad sensum spiritualia et subtilia declarabunt.
6     </quote>
7 </cit>

```

1. 17 Linking, Segmentation, and Alignment - The TEI Guidelines, URL : <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/SA.html> (visité le 15/08/2024).

Cependant, en plus de ce changement dans le XML, il aurait également fallu modifier le code de la base de données pour faire apparaître les liens. Cela aurait aussi posé problème car cela modifiait la structure globale du XML. Par conséquent, avec l'ingénieur de recherche Emmanuelle Khury, nous avons décidé d'intégrer les liens dans une balise préexistante pour éviter tout changement dans le XML. Nous avons choisi de placer le lien dans l'attribut correspondant de la balise `<ref>` qui se trouve dans la balise `<bibl>`. Cette dernière est utilisée dans SourcEncyMe pour indiquer une œuvre, comme vous pouvez le voir ici :

```

1   <cit n="2" xml:id="cit_id394697881596">
2     <bibl>
3       <ref target="#historiarum_adversum_paganos_etrosii_libri_vii" type="oeuvre">Historiarum adversum paganos Orosii libri VII</ref>
4       <author ref="#etrosius_paulus">Orosius Paulus</author>
5     </bibl>
6     <bibl>
7       <ref cert="low" corresp="https://thema.huma-num.fr/exempla/TE022184" target="#exemplum">Exemplum</ref>
8     </bibl>
9     <quote>
10       <seg type="marqueur">Orosius de hormesta mundi libro primo</seg>Extant adhuc certissima monumenta gestorum. Nam tractus curruum
11         et orbite rotarum, non solum in littore, sed etiam in profundo, nunc quoque visus admittitur pervidentur. Et si forte
12         ad tempus casu vel curiositate conturbatur, mox divinitus in pristinam faciem ventis, fluctibusque reparantur, ut
           quisquis non docetur timore dei vel probate religionis studio, ita eius transacte ultionis terreatur <seg type="marqueur"
           ">exemplum.</seg>
13     </quote>
14   </cit>

```

De plus, un attribut `cert="low"` a été ajouté pour indiquer un faible taux de certitude, étant donné que l'ajout des liens a été automatisé². De plus, si plusieurs liens sont présents, ils seront séparés par des espaces dans le même attribut `corresp`. Le lien sera visible car ce qui est inclus dans les balises `<bibl>` des balises `<cit>` est déjà affiché par le code de la base de données :

Ceci est une normalisation de la référence médiévale renvoyant au mémento de l'œuvre et de l'auteur, non une identification de source. Pour voir l'identification de la source (si elle existe déjà), cliquer sur la citation.

Auteur	Oeuvre	Memento
Hugo de Folieto	De clastro anime	Auteur allégué : Hugo de Folieto Auteur réel : Hugo de Folieto Titre : De clastro anime

[Hugo de Folieto] [ex primo libro de clastro anime] Audient seculares quos delectatio temporalium retinet, qui religiosi abhorrent habitum quam suavis sit, vere religionis cohabitare. Habundans enim est pauperi, mediocri sufficiens, tolerabilis dixit, infirmis larga, delictis compatiens, fontonibus moderata, penitentibus misericors, perversis severa, bonis optima. Hec sunt IX beneficia religionis. Quidam rebus temporalibus egent, sed inviti et hec vocari solet egestas, egent enim rebus et egent bona voluntate quia datus nichil esse potest. Quidam secundo veniunt ut in ecclesia possint honorari, qui in suo domo non nisi contemptibiles esse potuerunt. Qui si forte ecclesiam in qua paupertas amatuer intraverint, intolerabiliem dicunt illius loci consuetudinem, nimiam prelatorum severitatem pretendunt. Aliquem de fratribus opponunt sibi quem cum nolint pati dicunt quod non possunt, collide machinantes at ditionem valeant transire ecclesiam quasi humiliter rogant a vinculo professionis absolviri, miscentes blandis aspera. Minutur enim scandalum, si male discesserint et iurant in alia ecclesia se quiete mentis posse esse, in sua vero nunquam. Respondent abbati suo fronte liber, accusant fratres quosdam criminiosos quosdam per yroniam iustos vocant, quos vero iustos esse negare non possunt, inadvertentes bono nimis iustos appellari. Genus suum erubescunt, proprio nomine nolunt appellari. Amant enim nomina dignitatum, si vocarent abbates, si priores, si prepositi, arridet oculus, hylarescit facies. Et sic nunciat animus quid optet. Huiusmodi pauperibus non promittit **Christus** regnum, sed infernum. Sunt ali qui ex paupertate venerunt, quibus si dentur necessaria ex humiliitate putant esse superflua. Ocum dici unius sacramentum esse credunt, saturi vilibus cibis arbitrantur vitum, nulli equales esse, ex debito cunctis servire estimant, aliquando bonas erubescunt vestes habere, et nisi condiret humiliatatem discretio, ut licet licitum esse dubitarent. Tales sequuntur **Christum** pauperem itinere recto quia pauperes pauperem, humiles humilium. His datur ad necessitatem temporalia et promittuntur ad eternitatem celestia. Mediocri religio sufficere debet quia necessaria prebet, dat satis et dat gratis. Mediocritas est via media que dicit ad civitatem summi regis. A dextris vero et a sinistris sunt egestas et divicie. In una angustantur per egestatem, in alia vero errant per superbiam. In preciosa enim veste superbiam, in pannositate vero aliena rei concupiscentia timetur, mediocritas vero est quasi mensura qua totius bonitatis spacium terminatur.

FIGURE 2.10 – Affichage de l'œuvre dans SourcEncyMe

2. Il s'agissait de bien différencier les repérages faits de manière manuelle et ceux réalisés de manière automatique pour les chercheurs, dans un souci de transparence. Cette idée a été proposée par Emmanuelle Kuhry et Isabelle Draelants.

Enfin, pour indiquer la présence d’un *exemplum* dans le texte, des marqueurs sont utilisés pour surligner le début du récit exemplaire ou son introduction (voir le code au-dessus). Bien que ces marqueurs soient normalement utilisés pour signaler le début d’une citation, ils ont été détournés de leur fonction initiale pour les récits exemplaires³. Cependant, ces marqueurs ne sont présents que lorsque l’*exemplum* est explicitement introduit, ce qui n’est pas toujours le cas.

3. *SourceEncyMe...*

Chapitre 3

Repérages et recherches des données pour réaliser la connexion

Avant de créer les liens entre les *exempla* et les encyclopédies, il était essentiel d'identifier précisément les citations d'encyclopédies qui contiennent des *exempla* et les récits exemplaires qui s'appuient sur ces encyclopédies. Pour cela, j'ai ajouté un champ de recherche spécifique dans ThEMA et un nom d'autorité a été ajouté dans SourcEncyMe par Emmanuelle Kuhry. J'ai également envisagé de comparer les textes entre ThEMA et SourcEncyMe pour repérer les réemplois d'encyclopédies dans SourcEncyMe et les *exempla* dans ThEMA. En fin de compte, le repérage s'est principalement appuyé sur des mots-clés signalant les *exempla* dans SourcEncyMe et sur la présence de mentions d'encyclopédies déjà indexées dans ThEMA.

I. Crédit d'outils pour chercher dans les données qui seront connectées

Pour faciliter l'accès des utilisateurs aux informations ayant aidé à faire des liens, des outils spécifiques dans les deux bases de données ont été mis en place.

Dans SourcEncyMe, tous les *exempla* ont été catégorisés sous un même nom d'autorité (*Exemplum*) par Emmanuelle Kuhry, même s'ils ne peuvent pas être entièrement considérés comme tels car chaque récits exemplaire peut être considéré comme œuvre à part entière¹. Ce choix a été guidé par des contraintes techniques : toutes les sources de la base sont balisées de manière uniforme, et pour éviter de créer un nouveau système de balisage, le balisage existant a été adapté pour inclure les *exempla*. Ainsi, contrairement à ThEMA, où une distinction est faite entre différents types de récits, SourcEncyMe utilise

1. Cette autorité est désignée par un titre standardisé appelé entité canonique. *Ibid.*

le terme générique « exemplum » pour tous les récits exemplaires².

```

1 <cit xml:id="cit_idp94946128" n="11">
2   <bibl>
3     <ref target="#exemplum">Exemplum</ref>
4   </bibl>
5   <quote>Item activa vita assimilatur colli, primo, propter ordinem
prioratus : est enim collis pes montis. Nam per collum ascendimus ad montem, quia scilicet per vitam activam ad contemplativam venimus.
6       <name type="personne">Iacob</name> enim post <name type="personne">Lie</name> connubium (per quam activa vita signatur) ad <name
type="personne">Rachaelis</name> pervenit amplexum, per quam contemplativa vite formositas figuratur.</quote>
7 </cit>
```

Grâce à ce système, les utilisateurs peuvent avec un champ de recherche spécifique, qui existait déjà avant mon stage, rechercher tous les *exempla* présents dans les encyclopédies en saisissant « exemplum » dans le champ « Œuvre » :



FIGURE 3.1 – Champ de recherche dans SourcEncyMe

Dans ThEMA, en revanche, il a été nécessaire de créer un champ de recherche spécifique pour permettre aux utilisateurs d'explorer les sources et les textes apparentés. Bien que ces informations soient déjà visibles, il n'était pas possible de les rechercher directement. Pour remédier à cela, j'ai modifié le fichier « search.xql » situé dans le dossier « modules » de la base de données. J'ai adapté le code de recherche en texte intégral, initialement prévu pour les résumés des récits exemplaires, afin de permettre la recherche dans les sources et les textes apparentés (annexe F).

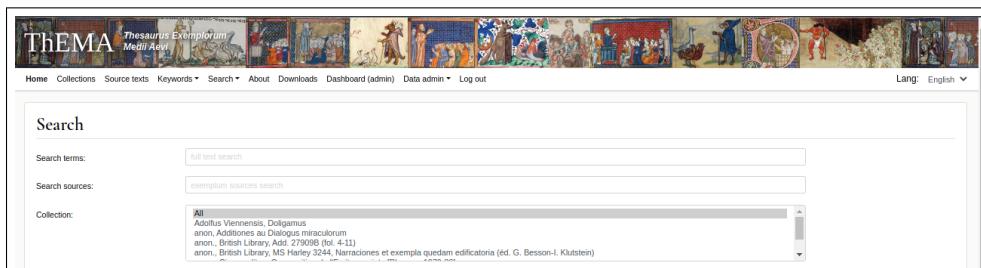


FIGURE 3.2 – Nouveau champ de recherche dans ThEMA

2. Informations sur ThEMA...

Toutefois, le développement de ce code n'a pas pu être finalisé avant la fin de mon stage. En l'état, le code affiche l'ensemble des *exempla* en haut de la page ainsi que les résultats de recherche dans les sources et textes apparentés en bas, alors qu'il aurait dû se limiter à afficher uniquement ces derniers.

The screenshot shows the ThEMA website interface. At the top, there is a navigation bar with links: Accueil, Recueils, Textes, Mots-clés, Chercher, À propos, Téléchargements, and Se connecter. A language dropdown menu shows "Lang: français". Below the navigation is a decorative header featuring a medieval manuscript illustration. The main content area displays search results for the query "exempla". There are two rows of results, each with columns for ID, Title, Author, and Description. The first row includes a "Mots-clés" column with links to related terms like "Femme", "Mort", "Chevalier", etc. Below the results is a pagination bar showing "Showing 1 to 50 of 5,469 entries" and a "Previous" button followed by a page number "1" which is highlighted in blue, with other numbers 2, 3, 4, 5, ..., 110, and "Next".

ID (rec.)	ID (ex.)	Auteur	Titre	Exempla	Mots-clés
TC0001	TE001278	Thomas Cantipratensis	Bonum universale de apibus:	Une femme voit monter au ciel l'âme (sous la forme d'un bel oiseau) d'un chevalier mourant après la bataille de Walcheren.	Femme Mort Chevalier Oiseau Âme Guerre Odeur
TC0001	TE001262	Thomas Cantipratensis	Bonum universale de apibus:	Une pie domestiquée est mangée par un serviteur indélicat. Interrogée par Jean de Vicence, elle parle depuis le ventre du voleur.	Oiseau Miracle vol
Showing 1 to 50 of 5,469 entries					Previous 1 2 3 4 5 ... 110 Next
ID (rec.)	ID (ex.)	Auteur	Titre	Exempla	Mots-clés
TC0004	TE002711	Jordanus de Pisis	Giordano da Pisa, Vitae Patrum V, 39 (Patr. lat. 73, 895); Grégoire, Dialogi, III, 7 (Patr. lat. 77, 229); Jacobus de Voragine, Legenda aurea, [ed. Esempi [Balducciari, Graesse, 1890], cap. 137, p. 610; Vincentius Belvacensis, Speculum historiale [Douai, 1624], XXII, 73. 1993]; 81 (3)		Juif Démon Saint Signe croix
TC0004	TE002752	Jordanus de Pisis	Giordano da Pisa, Vincentius Belvacensis, Speculum historiale [Douai, 1624], IV, 64; Petrarca, De viribus illustris par. 34-45 [ed. Martellotti p. 66-67]; Gualterius Castillonensis, Alexandre sive Gesta Alexandri Magni libri X, Argumentum libri decimi, [Patr. lat. 209], 563-566; Il Boezio e l'Amighetto nelle versioni del Trecento [Battaglia, s.d.] 240; Trattati religiosi e libro degli exempli in antico dialetto veneziano [ed. G. Ulrich], 20; L'Intelligenza [Di Benedetto] [ed. Mistruzzì, 1928], p. 206.	Joie Alexandre le Grand douleur poison	
TC0004	TE002732	Jordanus de Pisis	Giordano da Pisa, Gregorius Magnus, Dialogi, 3, 7; Vincentius Belvacensis, Speculum historiale [Douai, 1624], 23, 73. Esempi [Balducciari, 1993]; 100 (2)		Juif Démon croix

FIGURE 3.3 – Résultat du nouveau champ de recherche dans ThEMA

II. Identification des liens par comparaison des textes

Après avoir mis en place ces outils pour faciliter la recherche d'*exempla*⁴ et d'encyclopédies dans ThEMA et SourcEncyMe, j'ai également exploré la possibilité d'utiliser des méthodes plus avancées pour affiner encore davantage cette analyse et trouver des liens qui n'ont pas encore été repérés. C'est ainsi que j'ai envisagé l'utilisation de l'intelligence artificielle pour comparer les textes des deux bases de données. L'idée était de tirer parti de la rapidité et de la précision de l'IA pour détecter non seulement les *exempla* dans SourcEncyMe, mais aussi pour identifier des mentions d'encyclopédies supplémentaires dans ThEMA, au-delà de celles déjà repérées lors de l'indexation initiale.

Toutefois, cette approche s'est rapidement heurtée à plusieurs défis. Le premier obstacle provenait d'une différence importante entre les deux bases de données. ThEMA, pour des raisons de droits d'auteur, ne peut fournir que de courts résumés en français des récits exemplaires, tandis que SourcEncyMe donne accès aux textes complets en latin, car des éditions récentes n'existent pas forcément. Cette disparité linguistique a rendu l'utilisation de l'intelligence artificielle complexe, nécessitant la traduction des textes des encyclopédies en français pour permettre une comparaison efficace³.

3. Je n'ai pas trouvé de modèles de langues entraînés pour le latin. De toute manière, ThEMA ne possèdent qu'un nombre limité de récits où le texte latin est indiqué

Traduction WebTran

Je vis aussi un autre bienheureux du même ordre, avancé à un âge très avancé, nommé Dodonus. Une nation frisonne, qui était diligente dans sa prédication, fit tant de progrès dans sa nation frisonne, qu'il la modéra grandement par sa férocité. Car dès les premiers temps, les Frisons avaient une coutume des plus cruelles : lorsqu'un homme d'une famille était tué par un autre, le corps tué n'était pas enterré dans la tombe, mais suspendu dans un cercueil, enfermé et séché dans la maison, jusqu'à ce que ceux de la famille opposée soient tués par vengeance, plusieurs ou au moins un des parents opposés le massacraient au lieu de le tuer par procuration : puis il livrerait d'abord ses morts à leur enterrement dû avec une grande solennité. Ledit frère a supprimé cette coutume la plus cruelle et la plus inouïe dans cette nation, et l'a promue à un état plus doux par de fréquentes exhortations.

Google Traduction

Je vis aussi un autre bienheureux du même ordre, avancé à un âge très avancé, nommé Dodonus. Une nation frisonne, qui était diligente dans sa prédication, fit tant de progrès dans sa nation frisonne, qu'il la modéra grandement par sa férocité. Car dès les premiers temps, les Frisons avaient une coutume des plus cruelles : lorsqu'un homme d'une famille était tué par un autre, le corps tué n'était pas enterré dans la tombe, mais était suspendu dans un cercueil, enfermé et séché dans la maison, jusqu'à ce que ceux de la famille opposée soient tués par vengeance, plusieurs ou au moins un des parents opposés le massacraient au lieu de le tuer par procuration : puis il livrerait d'abord ses morts à leur enterrement dû avec une grande solennité. Ledit frère a supprimé cette coutume la plus cruelle et la plus inouïe dans cette nation, et l'a promue à un état plus doux par de fréquentes exhortations.

Traduction ChatGPT

J'ai également vu un autre homme du même ordre, un homme bienheureux, très avancé en âge, nommé Dodon. Il était Frison de nation et, par sa prédication assidue, il fit tant de progrès parmi son peuple, les Frisons, qu'il atténuait grandement leur férocité. En effet, depuis les temps les plus anciens, les Frisons avaient cette coutume des plus inhumaines : lorsqu'un homme d'une famille était tué par une autre, le corps du défunt n'était pas enterré par les siens, mais conservé suspendu dans un cercueil et desséché dans la maison, jusqu'à ce que la famille opposée, pour venger le mort, en tue plusieurs, ou au moins un, en représailles. Ce n'est qu'alors que les Frisons confiaient leur mort à une sépulture digne avec une grande solennité. Ce frère, Dodon, par ses fréquentes exhortations, élimina cette coutume cruelle et inouïe chez ce peuple et les amena à un état plus clément.

Traduction Henri Platelle

J'ai vu un autre religieux du même ordre, un saint homme, très avancé en âge, du nom de Dodon et de nationalité frisonne, qui réussit par sa prédication inlassable à adoucir notablement la férocité de son peuple. Depuis des temps très anciens, les Frisons observaient en effet l'abominable coutume suivante : quand une famille avait perdu un de ses membres assassiné par quelqu'un d'une autre famille, le corps de l'assassiné n'était pas enseveli par les siens, mais il était gardé dans une pièce et il se desséchait dans sa maison jusqu'au moment où la famille lésée tuait plusieurs membres ou tout au moins un membre de la famille adverse pour venger l'assassiné par une mort de remplacement. C'est alors seulement qu'elle enterrait solennellement son mort dans une sépulture normale. Le susdit frère fit abandonner par ce peuple cette coutume cruelle et inouïe et par ses fréquentes exhortations il l'amena à des habitudes plus douces.

FIGURE 3.4 – Différentes traductions d'un même *exemplum*

La traduction elle-même a posé des problèmes. Malgré plusieurs tentatives, je n'ai trouvé aucun traducteur en ligne parfaitement fiable. Parmi les outils testés, ChatGPT se distinguait légèrement de Google Traduction et de WebTran, mais ces outils imposaient des restrictions sur le web scraping, la technique d'extraction automatisée de données à partir de sites web. Pour ChatGPT, l'achat de jetons était requis pour accomplir cette tâche, tandis que Google Traduction et WebTran interdisaient explicitement le web scraping dans leurs conditions d'utilisation, compliquant encore davantage le processus.

En fin de compte, la solution retenue a été de traduire manuellement, ou partiellement, les textes des encyclopédies en m'appuyant sur des ressources en ligne. Bien que la traduction reste imparfaite, un code que j'ai développé (annexe G) a permis de repérer certaines similitudes entre les textes. Cependant, même avec des optimisations, ce traitement s'est révélé extrêmement lent : chaque comparaison prenait environ 20 minutes, en raison du volume important de données à analyser.

Le code, bien qu'imparfait, s'est montré relativement efficace. Lors d'un test visant à trouver les nombreuses équivalences de l'*exemplum* de la « moniale chaste mais bavarde

»⁴, le code a réussi à identifier le texte utilisé pour la recherche avec une précision de 100 %, et les autres textes similaires se sont classés en tête des pourcentages de similarité :

	A Text	B Similarity
1	De même, même les lieux sacrés ne réjouissent pas les méchants. Et donc Grégoire : « Une moniale qui avait été très querelleuse tant qu'elle avait vécu, fut enterrée après sa mort dans l'église. La nuit suivante, le gardien de l'église eut la vision de la moniale coupée en deux devant l'autel. Une partie (de son corps) était consumée par le feu, l'autre demeurait intacte. Lorsque le gardien de l'église raconta ces faits le lendemain et désigna l'endroit où était la flamme, l'embrasement fut visible comme si cela avait été la combustion d'un feu matériel. »	1
2	Une religieuse, qui faisait preuve de continence mais s'avérait être également bavarde et agressive, meurt. Elle est ensevelie dans l'église. La nuit suivante, le gardien la voit traînée hors de son tombeau et coupée en deux devant l'autel. Une partie brûle, l'autre reste intacte. Le lendemain matin, il raconte ce qu'il a vu aux frères qui découvrent à l'endroit indiqué les traces d'un feu matériel.	0,88072681427002
3	Une moniale chaste mais bavarde, impudente et futile fut enterrée dans l'église. La nuit même, le gardien eut la révélation qu'une partie du corps brûlait, l'autre restant intacte : des flammes apparaissaient en effet devant l'autel, marquant ainsi la matérialité du feu autour de la morte.	0,842399120330811
4	UNE MONIALE CHASTE MAIS QUERELLEUSE, BRULÉE. Le gardien d'une église vit une moniale chaste mais querelleuse, insolente et bavarde, conduite, pendant la nuit devant l'autel et brûlée sur la moitié du corps.	0,827393114566803

FIGURE 3.5 – Essai de comparaison

Malgré ces résultats encourageants, j'ai finalement décidé de renoncer à l'utilisation de l'IA pour trouver des correspondances entre les textes des deux bases de données. Cette décision s'explique par les nombreuses limitations rencontrées, notamment les problèmes de traduction, la lenteur des analyses, et la difficulté à définir ce qu'est un *exemplum* : une forme littéraire complexe et fluide, difficile à cerner tant par un humain que par un algorithme⁵. De plus, les *exempla* présents dans les encyclopédies diffèrent souvent de ceux répertoriés dans ThEMA, rendant la tâche encore plus ardue pour une IA. Un exemple illustrant cette difficulté est un *exemplum* retrouvé dans la *Summa de exemplis ac similitudinibus rerum* de Giovanni da San Gimignano, où l'auteur reprend une idée d'Aristote sur la nécessité de cultiver son cœur pour produire de bons fruits⁶. Ce récit n'a trouvé aucun équivalent exact dans le corpus de ThEMA, le texte le plus proche présentant seulement 70% de similarité.

Car comme le dit Aristote, les arbres sauvages deviennent des arbres de jardin par une culture appropriée ; il donne l'exemple de l'amandier, qui, s'il est amer, devient doux grâce à la culture. De même, l'homme doit cultiver et purifier son cœur pour produire de bons fruits.

FIGURE 3.6 – *Exemplum* repéré dans SourcEncyMe

4. *Les récits exemplaires de la moniale chaste et bavarde dans ThEMA*, URL : <https://thema.huma-num.fr/search?term=&collection=all&author=all&context=all&keyword=&keyword=&keyword=&tubach=TUB723&tubach=&tubach=&allegory=all> (visité le 01/09/2024).

5. J. Berlioz et Marie Anne Polo de Beaulieu, « Introduction générale », dans *Le tonnerre des exemples : Exempla et médiation culturelle dans l'Occident médiéval*, dir. Marie-Anne Polo De Beaulieu et Pascal Collomb, Rennes, 2010, p. 11-15, DOI : 10.4000/books.pur.132003.

6. *Récit exemplaire utilisé pour la comparaison de textes*, URL : http://sourcencyme.irht.cnrs.fr/encyclopedie/summa_de_exemplis_ac_similitudinibus_rerum_ed_anvers_1609?citid=cit_idp119842416 (visité le 01/09/2024).

	A	B	C
1	Texte	Similarité	
2	On raconte que les arbres de la forêts vinrent voir l'olivier, le figuier et la vigne pour qu'ils deviennent roi des arbres.	0,73338449001312	
3	L'homme est comparé à un arbre à l'envers : il doit tourner ses fruits vers la vie présente, ses racines (cheveux) vers le ciel.	0,7062361240387	
4	L'arbre produit des feuilles, des pierres précieuses, des fleurs et des fruits : respectivement les symboles des bonnes paroles, des bonnes pensées, des bonnes volontés et des bonnes œuvres.	0,68076729774475	
5	Le roi des arbres. Les arbres veulent un roi : l'olivier, le figuier et la vigne refusent, la petite branche accepte, mais n'a rien à donner. Les quatre plantes représentent respectivement la miséricorde, la contemplation, la charité et l'ambition. <small>La dieu des arbres et la plante</small>	0,67146623134613	

FIGURE 3.7 – Comparaison de l'*exemplum* avec les récits exemplaires de ThEMA

Ainsi, j'ai préféré adopter une approche plus simple : repérer semi-manuellement les *exempla* dans les encyclopédies. Contrairement à ThEMA, où les liens avec les encyclopédies ont été établis lors de l'indexation, ces *exempla* n'ont pas encore été identifiés dans les encyclopédies. L'entraînement d'un modèle de langage pour reconnaître un *exemplum* aurait été une solution intéressante, mais il aurait fallu surmonter à nouveau les défis précédemment présentés.

III. Identification des *exempla* via des mots-clés

Face à ces défis, il est apparu que l'utilisation d'une méthode plus ciblée pourrait s'avérer plus efficace. C'est ainsi que j'ai opté pour l'application de mots clés spécifiques. Cette méthode repose sur une caractéristique fréquente des récits exemplaires : ils sont souvent introduits par des termes latins comme « *audivi* », « *legi* », « *memini* », « *verum* », « *dicitur* », « *narrat* », « *memini* » ou « *vidi* »⁷. Cependant, tous ces mots n'ont pas été retenus, car dans les œuvres encyclopédiques, ils sont souvent utilisés dans d'autres contextes, notamment pour citer des œuvres ou des auteurs.

Par conséquent, je me suis concentré sur les mots les plus directement liés aux *exempla* : « *exemplum* », « *exempla* », « *exemplo* », et « *exemplis* ». Cette approche par mots-clés m'a permis de contourner les difficultés liées à l'intelligence artificielle tout en offrant une méthode efficace pour identifier les *exempla* dans les textes encyclopédiques.

De ce fait, avec une feuille de style XSL, c'est-à-dire une page de code utilisée pour transformer et présenter des documents XML, j'ai créé un code qui, lorsqu'il rencontre

7. Bremond Claude, Schmitt Jean-Claude historien et Le Goff Jacques, *L'Exemplum*, Turnhout, 1982.

l'une des déclinaisons d'*exemplum* dans la balise <quote> du XML (c'est-à-dire là où se trouve le texte), balise le mot avec un marqueur et crée ensuite une balise <bibl>, si nécessaire, au-dessus de <quote> pour indiquer qu'il y a un *exemplum*. Voilà ce que fait la transformation :

```

1   <cit n="7" xml:id="cit_idp82718352">
2     <quote>Sexto, quia malus seruus cupidus existens, bona Domini in proprium usum vertit. Exemplum de seruo <name type="personne">Elisei</
3       name>, et hoc pertinet ad simoniacos, qui bona Domini, id est, dona Spiritus sancti, pro sua utilitate temporali emunt, et vendunt
4   </quote>
</cit>
```

```

1   <cit n="7" xml:id="cit_idp82718352">
2     <bibl>
3       <ref cert="low" target="#exemplum">Exemplum</ref>
4     </bibl>
5     <quote>Sexto, quia malus seruus cupidus existens, bona Domini in proprium usum vertit. <seg type="marqueur">Exemplum</seg> de seruo <
6       name type="personne">Elisei</name>, et hoc pertinet ad simoniacos, qui bona Domini, id est, dona Spiritus sancti, pro sua
7       utilitate temporali emunt, et vendunt.
</quote>
</cit>
```

Au départ, je souhaitais que ce code fonctionne uniquement lorsque les déclinaisons du mot « exemplum » se trouvent en début de phrase, car j'avais remarqué que c'est souvent dans ces cas-là que l'on a affaire à un véritable *exemplum*. Cependant, après discussion avec Isabelle Draelants et Emmanuelle Kuhry, cette approche a été abandonnée car elle risquait de laisser de côté certains récits exemplaires. Le choix a été fait d'effectuer un repérage plus large, quitte à avoir plus d'erreurs (annexe H), ce qui a nécessité un travail de vérification.

Prenons l'exemple de l'encyclopédie de Giovanni da San Gimignano. Mara Calloni avait initialement identifié 220 *exempla* en lisant le texte. Le code, pour sa part, en a détecté plus de 650. Après vérification par mes soins, le nombre exact d'*exempla* est de 479. Bien que ce chiffre puisse sembler élevé, il est important de noter que l'ensemble de l'œuvre encyclopédique de San Gimignano est moralisée⁸. Le code repère donc un plus grand nombre d'*exempla* que la lecture humaine, mais il se concentre uniquement sur les déclinaisons du mot *exemplum*. En revanche, le repérage manuel effectué par Mara, qui ne faisait aucune discrimination dans les citations, a permis de déceler des *exempla* même lorsqu'ils ne sont pas explicitement introduits par ces déclinaisons. Toutefois, l'ajout de mots introductifs supplémentaires pour affiner la recherche augmenterait considérablement la charge de travail pour la vérification, qui prenait déjà une quinzaine d'heures rien que pour les déclinaisons du mot *exemplum*. Il est donc essentiel de déterminer jusqu'où nous sommes prêts à aller dans ce processus de vérification, car autrement, cela reviendrait presque à effectuer une recherche par lecture. Cela dit, cette méthode facilite le balisage, puisqu'il suffit ensuite de supprimer les éléments qui ne correspondent pas à

8. Massimo Oldoni, « Giovanni da San Gimignano », dans *L'enciclopedismo medievale*, dir. Michelangelo Picone, Ravenna, 1994, p. 213-228.

un *exemplum*.

C'est surtout avec le mot *exemplum* que le plus grand nombre de récits exemplaires ont été trouvés (140), suivi par *exemplo* (30), *exempla* (13) et *exemplis* (5). Cette répartition s'explique aisément. *Exemplum* (nominatif singulier) est la forme de base du mot en latin⁹. Elle est souvent utilisée pour introduire un récit exemplaire en tant que concept ou titre. *Exemplo* (ablatif ou datif singulier) est fréquemment utilisé dans les constructions syntaxiques pour indiquer le moyen ou la cause (par exemple, « par un exemple », « grâce à l'exemple de... »)¹⁰. Cette forme est souvent intégrée dans des phrases qui illustrent ou justifient un point par l'utilisation d'un exemple, ce qui la rend très présente dans les textes qui visent à enseigner ou moraliser.

9. Page dédiée au mot *exemplum* dans Wiktionnaire, juin 2024, URL : <https://fr.wiktionary.org/w/index.php?title=exemplum&oldid=35038953> (visité le 02/09/2024).

10. *Ibid.*

Chapitre 4

Création des liens dans les bases de données

Après avoir identifié les *exempla* dans SourcEncyMe et les encyclopédies dans ThEMA, il a été nécessaire de créer des connexions entre les deux bases de données. Pour ce faire, j'ai d'abord indexé les *exempla* identifiés dans SourcEncyMe en les regroupant dans un recueil factice où les liens pouvaient être ajoutés. Ensuite, j'ai développé un code permettant d'ajouter des liens vers les *exempla* qui contiennent des références à des encyclopédies.

I. Indexation des *exempla* de SourcEncyMe dans ThEMA

Une fois les récits exemplaires repérés dans SourcEncyMe, il était essentiel de les connecter à ThEMA. La question s'est posée de savoir où et comment les placer dans la base de données. L'équipe de SourcEncyMe envisageait de créer des entrées d'*exempla* directement dans ThEMA en y ajoutant les informations pertinentes. Cependant, cette approche ne correspondait pas au fonctionnement de ThEMA, où les *exempla* sont toujours intégrés dans un recueil spécifique, respectant ainsi le contexte et l'auteur d'origine.

Avec l'accord d'Isabelle Draelants et de Marie-Anne Polo de Beaulieu, nous avons décidé de créer un recueil intitulé « *Exempla* de [Nom de l'encyclopédie] ». Cette solution s'inscrivait dans la vision de ThEMA, qui se veut une base de données d'œuvres contenant des *exempla*, sans se limiter aux seuls recueils d'*exempla*. De plus, lors de l'indexation, nous avons choisi d'inclure un extrait du texte en latin de l'*exemplum* trouvé dans l'encyclopédie dans le champ : « *texte original* ». Cependant, étant donné que le texte des encyclopédies est encore en cours de révision, il a été convenu que ces extraits seraient éventuellement supprimés pour éviter la diffusion de versions incorrectes. À l'avenir, des chercheurs ou des étudiants seront chargés de finaliser l'indexation en créant des résumés et en ajoutant des mots-clés, après quoi les extraits en latin seront retirés. L'indexateur pourra également

supprimer un *exemplum* s'il estime qu'il n'en est pas véritablement un, car il ne faut pas oublier qu'une partie des *exempla* a été identifiée automatiquement par un algorithme. Dans ce cas, il devra simplement contacter l'équipe de SourcEncyMe pour qu'ils mettent à jour le lien de leur côté.

Pour intégrer les *exempla* dans ThEMA, j'avais initialement envisagé d'utiliser un code Python ou XQuery pour générer les fichiers XML des *exempla*. Ils auraient recréé le schéma XML des *exempla* de ThEMA et il aurait juste fallu réintégrer les XML des récits exemplaires dans la base pour qu'ils apparaissent. Mais j'ai finalement opté pour une solution différente. J'ai développé un script Python de web scraping (annexe I) qui tire parti de l'interface de création d'*exempla* déjà existante dans la base de données. Ce script se connecte à la base de données locale, navigue dans le recueil créé, crée chaque *exemplum*, y insère le texte source, et répète l'opération pour chaque *exemplum* repéré. Le code cible plus spécifiquement toute les <ref, target="#"#exemplum"> et récupère le texte dans la balise <quote> en dessous. Le code ne cible par le marqueur *exemplum* car il n'est pas toujours présent même si un *exemplum* est dans le texte. Cette méthode a l'avantage de créer directement les *exempla* dans la base, sans avoir à charger manuellement les fichiers XML, réduisant ainsi les risques d'erreurs.

Ce code pourra être réutilisé à l'avenir pour accélérer l'indexation des textes dans ThEMA, à condition que les *exempla* soient préalablement balisés. Il pourrait également être amélioré en intégrant un module capable de générer automatiquement des mots-clés pour les résumés, ce qui rendrait le processus d'indexation encore plus rapide.

II. Gestion des liens vers SourcEncyMe dans ThEMA

1. Crédation des liens

Lors de l'intégration des liens dans mon code, j'ai rencontré un problème. Le créateur d'*exempla* de la base ThEMA possède un champ spécifique pour ajouter ces liens, mais ce champ est « dynamique ». Cela signifie que la page peut se modifier ou réagir aux actions de l'utilisateur sans nécessiter un rechargeement complet. Cependant, mon code ne parvenait pas à interagir correctement avec cette partie dynamique, qui permet d'ajouter ou de supprimer des liens.



FIGURE 4.1 – Page dynamique pour ajouter des liens dans ThEMA

Finalement, il a été nécessaire de développer un script Python pour modifier les fichiers XML et y ajouter les liens (annexe J). Pour cela, j'ai dû télécharger le dossier contenant les *exempla* de Giovanni da San Gimignano. Le script vérifie d'abord que le texte dans les XML de ThEMA correspond bien à celui dans SourcEncyMe. Une fois cette vérification effectuée, le lien vers SourcEncyMe est ajouté dans les XML de ThEMA, et dans les XML de SourcEncyMe. Les liens sont construits comme suit :

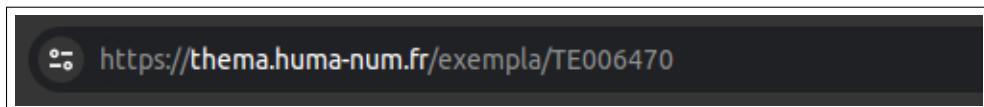


FIGURE 4.2 – Exemple de lien vers ThEMA

Pour accéder aux exempla, la base ThEMA passe d'abord par Huma-Num qui héberge la base. Ensuite, elle interroge le fichier « exempla.xql » pour afficher les *exempla* sous forme de contenu web. Enfin, elle appelle le fichier XML correspondant à l'*exemplum* créé dans ThEMA.

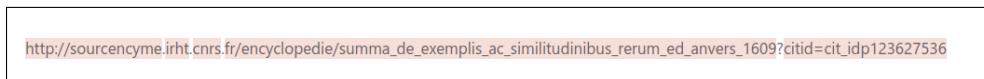


FIGURE 4.3 – Exemple de lien vers SourcEncyMe

Pour accéder aux citations d'encyclopédies, la base SourcEncyMe passe par l'IRHT et le CNRS, qui hébergent la base de données. Ensuite, elle interroge le fichier gérant les encyclopédies et leur affichage web. Enfin, elle appelle une encyclopédie spécifique et une citation précise identifiée par un code `cit_id`, unique pour chaque citation, trouvé dans la balise `<cit>` :

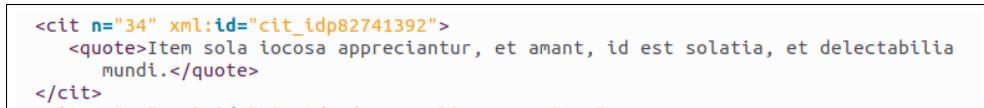


FIGURE 4.4 – Localisation du `cit_id` dans SourcEncyMe

Avant de créer les liens, j'ai vérifié auprès des deux équipes si les identifiants restaient constants dans le temps pour garantir la pérennité des liens.

2. Perspectives pour la création du lien

Dans une perspective d'amélioration future, j'ai envisagé de modifier le champ d'ajout de liens pour qu'il fonctionne de manière similaire aux autres champs contenant

des données préenregistrées. L’objectif serait de préenregistrer tous les liens vers SourcEncyMe, afin d’éviter aux chercheurs de les copier manuellement et de supprimer la nécessité de modifier directement les fichiers XML à l’aide d’un script Python.

J’avais envisagé d’apporter des modifications à l’interface de balisage XMLmind, utilisée pour baliser les encyclopédies de SourcEncyMe. Cependant, je n’ai pas eu le temps de m’en occuper, et c’est finalement Emmanuelle Kuhry qui s’est proposée de s’en charger. Il sera nécessaire d’adapter la fenêtre pop-up permettant de sélectionner une œuvre pour une citation dans XMLmind, en y ajoutant des liens vers les *exempla* de la base et la classification de Tubach.

III. Création de liens vers SourcEncyMe depuis les mentions d’encyclopedies dans ThEMA

Pour ajouter des liens entre les *exempla* et les références à des encyclopédies, j’ai utilisé un script Python (annexe K) qui explore les fichiers XML de la base téléchargée sur mon ordinateur. Le script recherche des mentions d’encyclopedies en utilisant une expression régulière spécifiée pour chaque encyclopédie. Par exemple, pour le *Speculum historiale*, l’expression régulière utilisée est :

```
r'S?s?peculum historiale,? ([IVXLCDM]+|\d+)[.,]?\\s*([0-9]+)?'
```

Cette expression régulière est conçue pour détecter les références au texte *Speculum historiale*, en tenant compte des variations possibles dans l’usage des majuscules, ainsi que des numéros de volume en chiffres romains ou arabes. Le script continue en récupérant les numéros de chapitre mentionnés et en cherchant ces numéros dans le fichier XML de l’encyclopedie. Une fois le numéro de chapitre trouvé, il est converti en chiffre arabe si nécessaire. Le script récupère ensuite le cit_id correspondant et l’insère dans un lien préétabli au nom de l’encyclopedie (à ajuster par l’utilisateur), puis intègre ce lien dans le XML de ThEMA. Finalement, le script place l’identifiant TE... dans le champ cit du XML, car il s’agit d’un élément repris et non d’un *exemplum*, ce qui évite l’identification d’une œuvre.

Cependant, plusieurs problèmes ont été rencontrés lors de l’exécution de ce code. Premièrement, l’extraction des mentions d’encyclopedies s’est révélée compliquée en raison de l’utilisation mixte d’indexations en texte intégral et d’attributs récupérés lors du chargement de la page web de la base de données. Les attributs spécifiques associés aux encyclopédies, stockés dans le fichier list_bibliography.xml, sont :

Z-S7UCXT48	<i>Speculum naturale</i>
Z-38N6WX3E	<i>Speculum historiale</i>
Z-D6KRQBM	<i>Liber de natura rerum</i>

Le script devait donc non seulement effectuer des recherches en texte intégral mais aussi intégrer et traduire ces codes avant de réaliser les recherches en texte intégral. Voilà à quoi ressemble un cas de mention plein texte d'encyclopédie dans un XML de ThEMA et une mention d'encyclopédie liée à un attribut :

```

1 <sourceDesc>
2   <list type="source_details">
3     <item type="source_text"><p></p></item>
4     <item type="sources"><p>Vincent de Beauvais, Speculum naturale, 16, 97 (Douai, 1624, col. 1213).</p></item>
5     <item type="exemplum_context"><p>III, tit. VI, De peregrinatione</p></item>
6     <item type="commentary"><p></p></item>
7     <item type="allegory">n</item>
8   </list>

```

```

1 <listBibl>
2   <bibl type="tubach" corresp="Z-WLZ7CBVC" n="TUB3969"/>
3   <bibl type="related_texts" corresp="" n="">Valerius Maximus, Dictorum factorumque memorabilium libri novem..., V, c.4.7.ext.</bibl>
4   <bibl type="related_texts" corresp="Z-38N6WX3E" n="6, 125"/>
5   <bibl type="related_texts" corresp="" n="">British Library, Add. 33956 [transcr. Welter], 427</bibl>
6   <bibl type="related_texts" corresp="" n="">Liber ad status [Paris, BnF, ms. lat. 6368], 4,21</bibl>

```

```

1 <biblStruct xml:id="Z-38N6WX3E" type="book" corresp="http://zotero.org/groups/2304628/items/38N6WX3E">
2   <monogr>
3     <title level="monograph">Speculum historiale</title>
4     <title level="short"/>
5     <author>
6       <name>Vincentius Belvacensis</name>
7     </author>
8     <imprint>
9       <pubPlace>Douai</pubPlace>
10      <date type="pub_date">1624</date>
11    </imprint>
12  </monogr>
13 </biblStruct>

```

Un autre problème rencontré concernait la différence entre les numéros de chapitres indexés dans ThEMA et ceux présents dans SourcEncyMe. Les numéros de chapitres utilisés par les indexateurs pour le *Speculum historiale* sont basés sur l'édition de 1624, tandis que le classement de SourcEncyMe est décalé d'un numéro¹. Ainsi, chaque fois que le code récupère un numéro de chapitre de ThEMA, il doit ajouter un à ce numéro pour correspondre à SourcEncyMe. Par exemple, pour l'*exemplum* 844 de la *Scala Coeli*, les indexateurs ont mentionné le chapitre 26, 32, alors que dans SourcEncyMe, ce texte se trouve au chapitre 27, 33.

1. Vincent de Beauvais, *Speculum historiale. Bibliotheca mundi Vincentii Burgundi, ex ordine Praedicatorum venerabilis episcopi Bellovacensis, Speculum quadruplex, naturale, doctrinale, morale, historiale*, Douai, ex officina typographica Baltazaris Belleri, 1624.

Summary:	(Fr.) Saint Jacques porte secours à un pèlerin abandonné par ses compagnons.
Sources:	Refert Calixtus... = Speculum Historiale, XXVI, 32.
Keywords:	saint James the Elder pilgrimage illness Death Horse burial Penance
<hr/>	
<h2>Original text</h2>	
<p>Item ad idem. Refert Calixtus papa quod cum .XXX. Loringi promisissent peregrinacionem beatu[m] Gobi, unus alium non dimitteret nec deficeret, et de hoc sibi invicem fidem dedisset. Cum unus eorum infirmaretur graviter et per .XV. dies eum portassent, aliqui affect fedio fidem frangentes dimiserunt eum, nec remansit cum eo nisi unus qui se non abstinuerat juramento. Cum autem de vita infirmi desperarentur eum portassent iste usque ad pedem montis, dum supra eum dederidisset, monaus est infirmus. Cujus horrem non valens stans, iste socius cum esset solus et terra esset periculosa, fendo oravit beatum [Gobum]. Quia statim ei apparens in peregrina forma cum equo, requisitus causam fletus. Qui cum expressisset, beatus [Gobus] accepit mortuum super equum ante se, et in spatio illius noctis per .XII. dietas eum portauit usque ad montem Gaudi, qui est prope Sanctum [Gobum] ad unum leucum. Et precepit canonicus ex parte beatu[m] Gobi ut sepelirent corpus portatum. Tunc dicti beatus [Gobus] socio fideli quem super equum retro se portauit ut diceret socios suis qui in civitate Longiorensi erant ut pertinetiam agerent de fide fracta, quia cum peccato nulla pernigracito placabilis est Deo. Et statim dispergit. Qui ad eos accedens, dum vidisset miraculum egerunt pernitemant agerent de fide fracta.</p>	
From:	POLO DE BEAULIEU/ Marie Anne (éd.), 1991. Jean Gobi. La Scala coéf., Paris, Editions du CNRS.
View in edition	

FIGURE 4.5 – Récit exemplaire reprenant un texte encyclopédique

```
<div n="32" subtype="chapitre" type="section" xml:id="speculum_historiale.27.32">
  <head>ADHUC DE XXXIII. </head>
  <cit n="1" xlink:href="cit_id394698275640">
    <bibl>
      <ref target="#miracula_sancti_iacobi_maioris" type="oeuvre">Miracula sancti Iacobi maioris</ref> <author
ref="#hubertus_bisuntinus">Hubertus Bisuntinus</author>
    </bibl>
    <quote>
      <seg type="marqueur">Hubertus bisuntinus</seg>Anno domini Mo LXXXO, XXX viri de-nome>Lotharingia</
name>sanctum<name>Iacobum</name>adeuntes mutue servandam virtutis fidem sibi huncem promiserunt. Unus tamen inter eos fuit, qui hoc eis
non promisit. Cum itaque ad urbem qua porta clausa dicitur incolumes pervenissent, unum e sociis ibi egrotantem propter pactum quod
fecerant per IV dies usque ad portas<name>Tiferenos</name>cum grandi labore tulerunt. Cum illi XV dieate ab expeditis V diebus agl possent,
tunc tandem gravati infirmum reliquerunt. Ille tamen solus qui fidem non promisit cum eo remansit ad pedem montis sancti<name>Michaelis</
name>. Tunc dixit infirmus ut eum si posset super illum montem ferret. Quo liberenter ille deuult et die declinante ad vesperam infirmus
migravit. Quod videns vivus valde timuit proximum imminentem caliginem et defuncti presentiam et gentis barbare fertatem. Sic ergo
omni modo destitutus ad orationem se contulit et statim sanctus<name>Iacobus</name>et in specie equitis apparsens causam fletus inquirit:
Domine, inquit, quia est nox et mortuus iste sepeliri non potest. Tunc illi: Trade micht hinc mortuum et tu ascede post me super equum.
Quo facta, nocte illa ante solis ortum XII dieatas peragentes ad-name>Montem</name>gaudii, qui est dividia leuca citra
sanctum<name>Iacobum</name>, pervenerunt. Ibique sanctus vivum et mortuorum de equo depositus, lubens vivo ut canonicos sancti<name>Iacobis</
name>ad sepielenund peregrinum invitarerat et peracta peregrinatione cum invenirent socios in urbe qua dicitur<name>Legio</name>, eos de
fracto pacto corriperent et dicseret eis ex parte sancti<name>Iacobis</name>, quod propter hoc eum peregrinatio non valereret, et his
dictis disparuit. Ille autem rediens, cum hoc inventis sociis nunclasset, valde miratus sunt et ab episcopo<name>Legionis</name>: civitatis
ubi ille consecratus fuerat, super hoc penitentiam suscepérunt.</quote>
  </bibl>
</div>
```

FIGURE 4.6 – Texte de l'encyclopédie qui est repris

Le dernier problème a résidé dans la division des encyclopédies en citations. Les liens peuvent uniquement être créés pour des citations spécifiques, alors que les numéros de chapitres indexés dans ThEMA ne fournissent que le numéro de chapitre, tandis qu'il peut y avoir plusieurs citations pour un même chapitre. Après discussion avec Emmanuelle Kuhry et Emmanuelle Kuhry, il a été décidé de placer les liens dans la première citation de chaque chapitre. Cette approche laisse à l'utilisateur le soin de trouver le passage exact. Cette solution a été retenue en raison de la complexité et du volume de vérifications manuelles nécessaires, étant donné que parfois un passage peut être mentionné dans toutes les citations d'un chapitre ou seulement dans une seule.

Chapitre 5

L’interconnexion des bases de données comme moteur de l’Open Data dans les sciences

Il est désormais important de remettre en perspective le travail de mise en relation réalisé ici. Ce travail s’inscrit dans une dynamique plus large d’ouverture des données de la recherche (Open Data), en contribuant à la fois à la collaboration scientifique, à la valorisation des données, et à la facilitation de leur accès pour les utilisateurs¹.

I. Renforcement des collaborations scientifiques

La connexion entre les deux bases de données a non seulement permis de relier des données éloignées, mais aussi de rapprocher des équipes de recherche géographiquement et institutionnellement distinctes².

Un rapprochement s'est opéré, notamment au niveau institutionnel. Au cours de ce stage, des réunions ont parfois été nécessaires entre les équipes des deux bases. Ces réunions avaient pour objectif non seulement de présenter l'avancement de mon travail, mais aussi de réfléchir collectivement aux défis rencontrés. Par exemple, une réunion avec Marie-Anne Polo de Beaulieu, Elisa Lonati et Isabelle Draelants a permis de tenir l'équipe de ThEMA informée de mes progrès alors que je collaborais avec l'équipe de SourcEncyMe. Ce moment a été crucial pour déterminer comment intégrer les récits exemplaires trouvés dans les encyclopédies au sein de ThEMA. Par ailleurs, chaque équipe a eu l'occasion de présenter son travail à l'autre avant le début du stage.

Ce rapprochement s'est également opéré au niveau des données, car j'ai réussi à connecter les deux bases qui, auparavant, étaient indépendantes. Cela a permis d'éviter

1. Bridgette Wessels, Kush Wadhwa et Rachel L. Finn, *Open Data and the Knowledge Society*, Amsterdam, 2017, doi : 10.1515/9789048529360.

2. Je me permets de renvoyer à l'introduction pour plus de détails

la duplication des efforts et de favoriser la réutilisation des données, tout en évitant que SourcEncyMe ne devienne une base spécialisée dans les exempla, et vice versa.

II. Valorisation des bases de données

La mise en relation des deux bases de données a également permis de valoriser chacune d’elles.

Ce stage m’a permis d’ajouter de la valeur à ces deux bases. Pour SourcEncyMe, j’ai contribué à l’identification des récits exemplaires dans les encyclopédies, qui n’avaient pas été repérés auparavant. Ce processus pourra d’ailleurs se poursuivre grâce au code XSL que j’ai développé. Pour ThEMA, j’ai enrichi la base avec de nouveaux *exempla* issus des encyclopédies, notamment en traitant intégralement deux œuvres : le *Speculum historiale* de Vincent de Beauvais et la *Summa de exemplis ac similitudinibus rerum* de Giovanni da San Gimignano. Il reste toutefois à télécharger les fichiers XML modifiés dans les deux bases pour rendre visible l’ensemble du travail réalisé durant mon stage³. Par ailleurs, les codes en annexes permettront de traiter toutes les autres encyclopédies de SourcEncyMe que je n’ai pas eu le temps de traiter pendant mon stage.

Ce stage a également contribué à rendre les deux bases de données plus accessibles et plus visibles. Les liens ajoutés augmentent la probabilité que les utilisateurs découvrent l’une des bases. J’ai aussi présenté, avec Marie-Anne Polo de Beaulieu, le travail en cours sur les deux bases de données au collectif « Sources et données de la recherche » du CRH, et un billet est prévu sur le blog de l’atelier de Vincent de Beauvais pour mettre en lumière ce travail. Par ailleurs, il est également envisagé d’informer le LabEx HaStec des avancées du travail.

III. Amélioration de l’expérience utilisateur

Enfin, la mise en relation des deux bases de données a facilité la recherche pour les utilisateurs. Auparavant, s’ils trouvaient un texte encyclopédique dans ThEMA ou un *exemplum* dans SourcEncyMe, ils ne savaient peut-être pas où chercher pour obtenir plus d’informations. Désormais, un utilisateur de SourcEncyMe qui trouve un *exemplum* peut consulter la bibliographie associée, découvrir d’autres récits exemplaires sur les mêmes thèmes, ou encore accéder à des textes similaires. De son côté, l’utilisateur de ThEMA pourra plus facilement retrouver le texte intégral de l’encyclopédie et identifier la source de l’auteur de l’encyclopédie.

Cependant, pour l’utilisateur, il s’agit souvent d’un « gain marginal » dans ce type de bases de données. En général, l’utilisateur explore une base pour une raison précise

3. Les fichiers ont été envoyés aux deux équipes, mais restent disponibles sur mon GitHub. <https://github.com/Laitauchocolat34/memoire-tnah>

et ne cherche pas nécessairement à exploiter toutes les possibilités offertes. Globalement, l’accessibilité en ligne de la base de données reste l’aspect le plus crucial, tandis que le reste peut être perçu comme un complément. Pour un utilisateur, la création et la mise en ligne d’une base de données sont plus importantes que l’établissement de liens entre différentes bases.

Chapitre 6

De nouvelles perspectives grâce à l’interconnexion

Il est désormais important de voir quelles perspectives offrent la connexion des deux bases données. Ce chapitre explore d’abord les avantages de cette interconnexion pour la recherche, puis présente les méthodes de visualisation qui permettent de clarifier ces connexions.

I. Bénéfices pour la recherche de la mise en relation des deux bases de données

La mise en relation des bases de données offre des avantages pour la recherche scientifique.

Pour le projet ThEMA, établir ce lien permettra d’approfondir l’étude déjà bien avancée des sources des récits exemplaires¹. Plus précisément, cela permettra de se concentrer sur les sources encyclopédiques, qui ont jusqu’à présent été peu analysées². En examinant ces sources, nous pourrons identifier les thèmes, les sujets, ainsi que les personnages historiques ou bibliques les plus souvent repris de manière globale dans les *exempla* depuis les encyclopédies. Les études des références encyclopédiques dans les recueils de récits exemplaires pourraient dépasser celles déjà réalisées par Marie-Anne Polo de Beaulieu et Jacques Berlioz sur la *Scala Coeli* et le *Tractatus de diversis materiis praedicabilibus*³. Nous pourrons aussi évaluer dans quelle mesure les récits exemplaires d’origine encyclo-

1. Il n’existe pas d’ouvrages offrant une synthèse des sources des récits exemplaires. Les sources diffèrent d’un recueil à l’autre, ce qui rend une telle synthèse irréalisable. Cependant, certaines éditions de sources ainsi que des articles ou chapitres de revues les analysent. La base bibliographique Bibliex, consacrée aux récits exemplaires, mentionne 36 fois le mot « source ». Le liens vers Bibliex : <https://www.zotero.org/groups/2304628/bibliex>

2. Il n’y a que 6 références aux encyclopédies dans Bibliex

3. J. Berlioz et M.A. Polo de Beaulieu, « Les recueils d’*exempla* et la diffusion de l’encyclopedisme médiéval »...

pédique ont été réutilisés par les compilateurs d'*exempla* ultérieurs. Les chercheurs auront l’occasion de comparer la façon dont divers auteurs ou œuvres utilisent les mêmes sources encyclopédiques dans leurs *exempla*, mettant en lumière des variations stylistiques, théologiques ou culturelles. En outre, cette recherche pourrait conduire à la découverte de nouveaux récits exemplaires, étant donné que les encyclopédies sont riches en textes inexplorés.

Côté SourcEncyMe, la mise en relation des deux bases pourrait éclairer les raisons pour lesquelles un auteur encyclopédique utilise un *exemplum*⁴. Cela permettra de comprendre pourquoi un auteur intègre un récit habituellement utilisé comme *exemplum* dans son encyclopédie, même si celui-ci ne sert pas nécessairement d'exemple. Cette démarche pourrait aussi révéler si des récits exemplaires issus de recueils d'*exempla* sont repris dans des encyclopédies et pour quelles raisons. Elle offrira la possibilité d'analyser plus en profondeur la place des récits exemplaires dans les œuvres encyclopédiques, ainsi que les raisons pour lesquelles ils sont insérés dans certaines sections d'une œuvre et non dans d'autres.

II. Techniques de visualisation pour les relations entre récits exemplaires et encyclopédies

Pour continuer la mise en relation des bases de données, plusieurs possibilités de visualisation des données pourraient être envisagées. Elles permettraient aux chercheurs et aux utilisateurs de mieux saisir visuellement les liens entre les encyclopédies et les récits exemplaires. Une pratique de plus en plus en plus utilisée par les chercheurs dans leurs publications⁵.

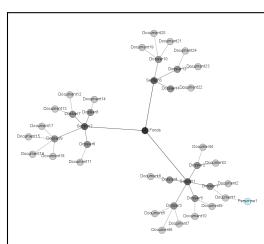


FIGURE 6.1 – Exemple de visualisation d'analyse de réseaux

4. Il n'existe pas d'étude spécifique sur l'utilisation des récits exemplaires dans les encyclopédies, bien que le lien entre encyclopédies et *exempla* ait déjà été établi, comme je l'ai mentionné dans l'introduction. La bibliographie sur les encyclopédies médiévales réalisée par Isabelle Draelants ne comporte que quelques rares mentions de récits exemplaires. Pour accéder à cette bibliographie, consultez : <https://ateliervdb.hypotheses.org/bibliographie-sur-lencyclopedisme-medieval>

5. Tianyin Jiang, Yixin Hou et Jaebum Yang, « Literature Review on the Development of Visualization Studies (2012–2022) », *Engineering Proceedings*, 38-1 (2023), p. 1-10, DOI : 10.3390/engproc2023038089.

Une approche intéressante serait la création de cartographies des liens établis par le travail de mise en relation des deux bases de données comme une sorte d’analyse de réseaux⁶. Par exemple, si plusieurs recueils d’*exempla* reprennent un même morceau d’encyclopédie ils seront connectés avec des traits vers un unique point matérialisant le texte de l’encyclopédie. Ces visualisations de réseaux pourraient permettre de représenter et de mieux saisir les emprunts et les influences entre les deux types de documents.

Une autre approche intéressante serait celle du graphe de citation c'est à dire une représentation visuelle des relations de citation entre différents éléments, dans ce cas, entre les récits exemplaires au sein des encyclopédies et les chapitres des encyclopédies. Le graphe de citations permet de voir rapidement quels récits exemplaires sont cités dans quels chapitres des encyclopédies. Chaque lien entre un récit exemplaire et un chapitre montre une relation de citation, c'est-à-dire que le chapitre utilise ou mentionne ce récit. Si un récit exemplaire apparaît dans plusieurs chapitres, il sera connecté à chacun d'eux par des liens distincts. Cela montre non seulement où le récit est utilisé, mais aussi l’étendue de son influence à travers l’encyclopédie.

Pour mettre en œuvre ces visualisations, il serait nécessaire d’adapter les fichiers XML selon les TEI Guidelines pour intégrer les métadonnées et annotations nécessaires⁷. Il faudrait ensuite avoir des logiciels comme Gephi⁸ ou Cytoscape⁹ pour développer les représentations. Pour ce qui est de l’analyse de réseaux il faudrait peut être un middleware car les deux bases de données restent séparées. Un middleware est un logiciel qui agit comme une passerelle entre des applications, outils et bases de données.

6. Laurent Beauguitte, « L’analyse de réseaux en sciences sociales et en histoire », dans *Le réseau. Usages d’une notion polysémique en sciences humaines et sociales*, dir. Rosemonde Letricot, Mario Cuxac, Maria Uzcategui Moncada et Andréa Cavaletto, Louvain-La-Neuve, 2016, p. 9-24, URL : <https://shs.hal.science/halshs-01476090> (visité le 31/08/2024).

7. *20 Graphs, Networks, and Trees - The TEI Guidelines*, URL : <https://www.tei-c.org/release/doc/tei-p5-doc/en/html/GD.html> (visité le 31/08/2024).

8. *Gephi - The Open Graph Viz Platform*, URL : <https://gephi.org/> (visité le 31/08/2024).

9. *Cytoscape : An Open Source Platform for Complex Network Analysis and Visualization*, URL : <https://cytoscape.org/> (visité le 31/08/2024).

Conclusion

En conclusion, j'ai réussi à relier deux bases de données XML-TEI en adoptant une approche techniquement réalisable, rapide, tout en maintenant une distinction institutionnelle claire. Pour ce faire, j'ai inséré semi-automatiquement des liens dans les fichiers XML des deux bases de données.

Cela a nécessité au préalable une étude approfondie des deux bases de données. J'ai dû m'assurer de la qualité et de la pérennité des données afin que les liens restent accessibles et durables dans le temps. J'ai opté pour la création de liens sur les données textuelles, en raison de leur utilité pour l'utilisateur et le chercheur mais aussi en raison du temps imparti. J'ai également vérifié où placer ces liens dans les XML et où trouver les informations nécessaires à leur création.

Ensuite, j'ai identifié précisément les textes à connecter. Cela a impliqué de rendre visibles les *exempla* dans SourcEncyMe et de permettre la recherche des sources et des textes apparentés dans ThEMA. J'ai exploré la possibilité d'utiliser l'intelligence artificielle pour comparer les textes, mais j'ai finalement choisi un repérage par mots-clés dans SourcEncyMe, car les références encyclopédiques étaient déjà mentionnées dans ThEMA.

La phase suivante a été la création des liens eux-mêmes. Pour cela, j'ai indexé automatiquement les *exempla* identifiés dans SourcEncyMe au sein de ThEMA. J'ai développé un code pour ajouter les liens dans ThEMA et SourcEncyMe pour les *exempla* associés à des encyclopédies. Ce code permettait également de récupérer les mentions d'encyclopédies présentes dans les *exempla* déjà indexés et de créer un lien chaque fois qu'une de ces encyclopédies était mentionnée dans SourcEncyMe.

Enfin, j'ai pris du recul sur le travail accompli pour démontrer qu'il contribue à un objectif plus large de la recherche, à savoir celui de l'Open Data. Ce projet a favorisé une plus grande collaboration scientifique et institutionnelle, valorisé les données, facilité l'accès des utilisateurs à ces informations, et ouvert de nouvelles perspectives pour la recherche sur les *exempla* et les encyclopédies.

Dans le contexte de l'Open Data et de l'interopérabilité, il est crucial de réfléchir aux liens vers d'autres bases de données dès leur création. Cela simplifie grandement le processus et réduit les problématiques liées à l'adaptation à différentes structures de bases de données.

Il ne faut pas pour autant abandonner les anciennes bases de données, même si elles semblent dépassées ou moins adaptées. Des personnes y ont investi du temps et des efforts, de l'argent et il est important de ne pas laisser ce travail se perdre, mais plutôt de les faire évoluer, comme l'a permis mon stage.

Annexe A

Installation de ThEMA

Étape 1 : Chargement de la base dans eXist-db

- Lancer eXist-db et ouvrir le localhost
- Se rendre dans l'onglet « Package Manager »
- Faire glisser le fichier « .xar » dans la « DROPZONE »
- Pour supprimer l'application ThEMA il suffit de cliquer sur l'icône poubelle dans la liste des packages installés

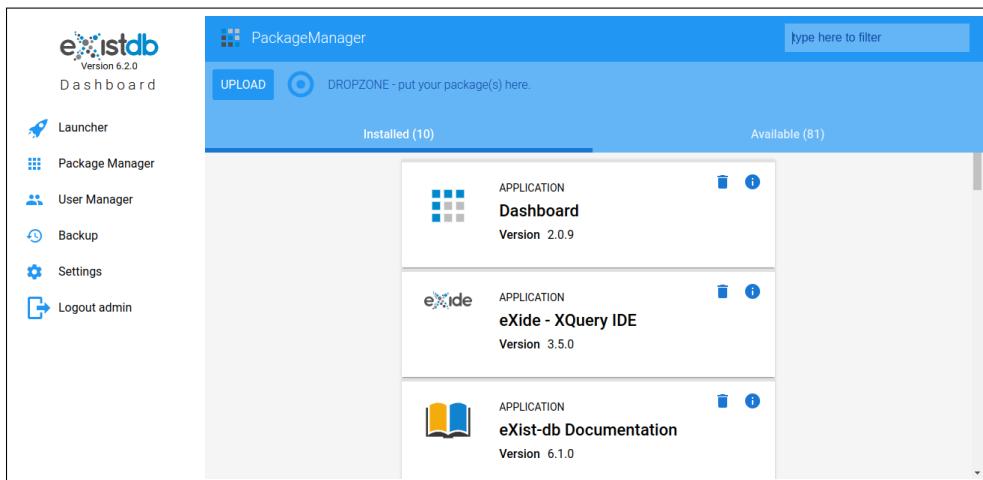


FIGURE A.1 – Page « Package Manager » de eXist-db

Étape 2 : Création des groupes de sécurité

- Faire clic droit sur l'icône eXist-db
- Cliquer sur Open Java Admin Client
- Créer un nom d'utilisateur, un mot de passe et mettre dans les favoris « localhost »
- Ouvrir « Outils » puis « Éditer les utilisateurs »
- Créer quatre groupes « indexer », « editor », « project-contributor » et « project-engineer »
- Par défaut, eXist-db nous identifiera comme admin

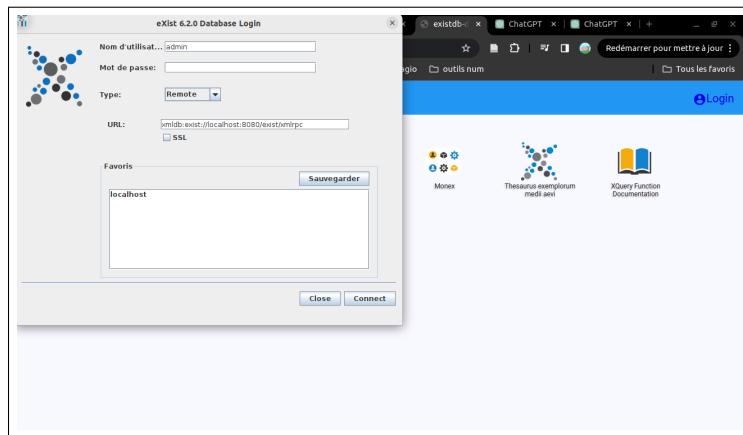


FIGURE A.2 – Modification dans l’Admin client n°1

- Aller dans « apps », « thema », « modules »
- Ouvrir « dbfunction » en cliquant sur « Editer les propriétaires »
- Cocher « SetUID » et « SetGID »
- Ouvrir « user.xql », cocher toutes les cases de la colonne « Execute » et passer du group « dba » à « thema »

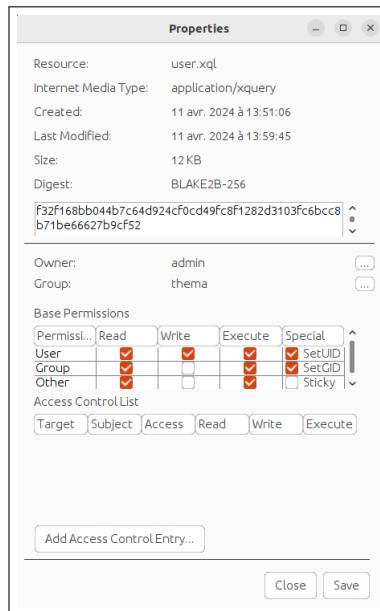


FIGURE A.3 – Modification dans l’Admin client n°2

Étape 3 : Compléter la base

- Aller dans eXide
- Aller dans « File » puis « Manage »
- Se rendre à l’endroit de l’arborescence et cliquer sur « Upload Files »
- Faire glisser les données et appuyer sur « done » quand le chargement est terminé

- Les dossiers et fichiers à ajouter sont ceux présents dans la copie de la base mais qui ne sont pas dans eXist-db
- Ne pas mettre trop de fichiers en même temps et bien attendre que la roue dentée arrête de tourner
- vérifier que les documents sont bien implémentés
- Se rendre dans « apps », « thema »
- Lancer les fichiers « .xconf » en appuyant sur « Eval »
- En cas d'ajouts de données, il sera nécessaire de lancer à nouveau l'évaluation
- Modifier « var base URL = » avec : « 'https://localhost:8080/exist/apps/thema/' »

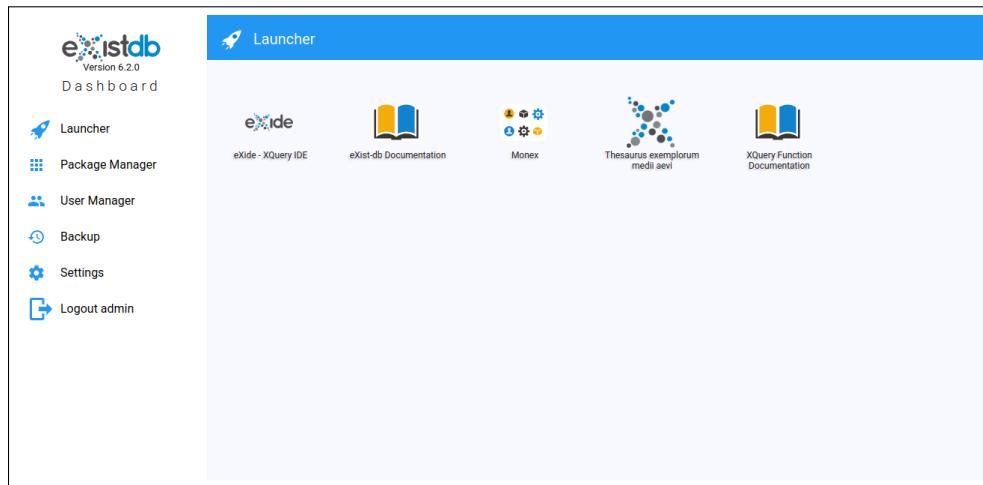


FIGURE A.4 – Page d'accueil d'eXist-db

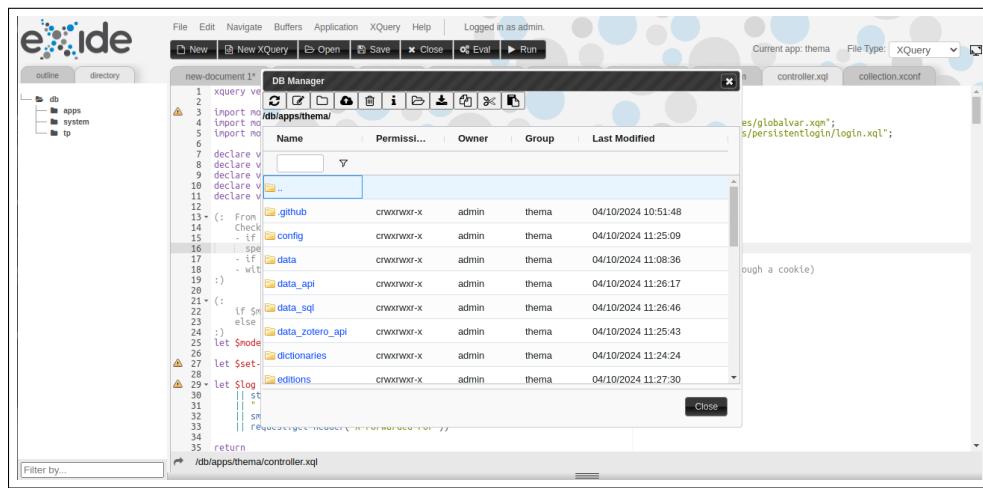


FIGURE A.5 – Page d'accueil « Upload Files »

Étape 4 : Lancer et éteindre la base de données

- Se rendre dans le « Launcher » de eXist-db
- Cliquer sur « ThEMA »

- Essayer toutes les fonctionnalités pour vérifier s'il n'y a pas de bug
- Cliquer sur l'icône eXist-db et cliquer sur « Stop server »
- Cliquer sur « Quit »
- Si cette manipulation n'est pas réalisée, il y a un risque d'altération des données et de la structure de la base

Étape 5 : Création d'un back-up

- Chercher l'endroit où est installé eXist-db sur l'ordinateur
- Ouvrir le fichier « conf.xml » et remplacer « true » par « false » dans « posix-chown-restricted »
- Ajouter dans les balises « <scheduler> </scheduler> », le code suivant :

```

1   <job type="system" class="org.exist.storage.ConsistencyCheckTask" cron-trigger="0 10 2 * * ?>
2     <parameter name="output-dir" value="backup"/>
3     <parameter name="zip" value="yes"/>
4     <parameter name="backup" value="yes"/>
5     <parameter name="incremental" value="no"/>
6     <parameter name="incremental-check" value="no"/>
7   </job>
```

- Aller dans « Fenêtre », « Afficher la vue », « Explorateur de sources de données »
- Se connecter à eXist-db localhost

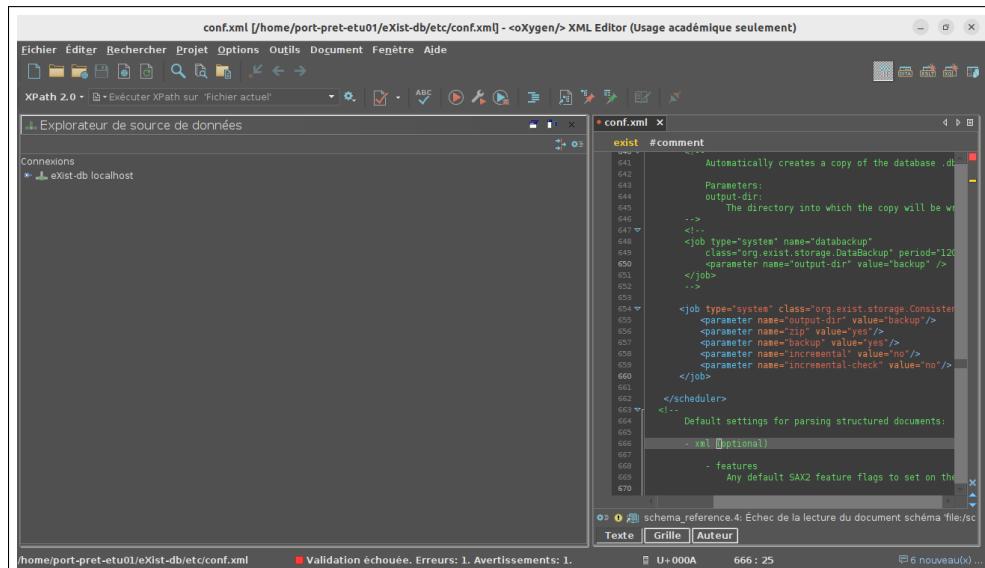


FIGURE A.6 – Utilisation de Oxygen pour coder en lien avec eXist-db

Annexe B

Fonctionnement de ThEMA

I. La structure de ThEMA

Dossier 1 : Data

- Les dossiers « TC » représentent les différents recueils d'*exempla*
- Les fichiers « TC » représentent l'écran d'accueil du recueil
- Les fichiers « TE » sont les différents *exempla* indexés
- Les fichiers sont listés par ordre d'ajout sur la base
- Liste des mots clefs
- Liste des manuscrits
- Liste de la bibliographie
- Liste des groupes de mots-clefs
- Liste des lieux
- Liste des personnes
- Liste des langues
- Liste des contextes
- Un fichier « xconf » sert à indexer les données dans la base une fois ajoutées

Dossier 2 : data_zotero_api

- Ce dossier contient l' de zotero.

Dossier 3 : editions

- Contient les éditions électroniques au format de quelques recueils d'*exempla*
- Contient aussi un fichier « .xconf » pour leur indexation dans la base

Dossier 4 : editorial

- Contient les informations éditoriales du projet ThEMA en 5 langues
- Contient un fichier avec les remerciements en 5 langues

Dossier 5 : modules

- Contient un dossier contenant le code pour générer des PDF - Contient tous les fichiers faisant tourner la base de données.
- Contient deux types de fichiers : xqm et xql

Dossier 6 : pdf

- Contient les présentations des recueils au format PDF.

Dossier 7 : ressources

- Contient un dossier « css » pour l'aspect visuel de l'application
- Contient un dossier « fontawesome » pour la police fontawesome
- Contient un dossier « fonts » pour les autres polices de caractères
- Contient un dossier « graphics » avec toutes les icônes
- Contient un dossier « images » pour les images de manuscrits présent dans la base
- Contient un dossier « js » pour le code en JavaScript
- Contient un dossier « json » pour le code json

Dossier 9 : temp

- Dossier vide

II. Le fonctionnement de ThEMA

Quand une requête arrive d'internet ou du localhost vers eXist-db, elle passe par Getty. Il s'agit d'une interface facilitant le dialogue entre internet et eXist-db. Pour coder ThEMA, il n'est pas nécessaire de consulter le code de Getty.

Le premier fichier important est controller.xql. C'est ici que la requête arrive après être passée par Getty. Il s'agit d'un échangeur qui va guider les requêtes vers différents endroits de la base de données. Il ne faut pas modifier ce code, sinon l'application ne fonctionnera plus.

Ensuite, la majorité du code se trouve dans le dossier modules. Le premier élément important dans ce dossier est le fichier globalvar.xqm. Ce code est un module qui définit plusieurs variables globales pour stocker des chemins d'accès, des URL et d'autres informations utiles. Globalement, il permet de créer des raccourcis.

L'autre module important est page.xqm. Il s'agit d'un module principal affichant la page d'accueil de la base dans un navigateur internet. C'est à partir de ce module qu'il est possible de naviguer sur le site de ThEMA. Les autres pages sont appelées avec « import module namespace ».

Le reste du site fonctionne comme « page.xqm ». En haut se trouvent les appels à d'autres pages, et en bas le code que le fichier traite. Il faut remonter progressivement les pages pour accéder aux données.

Pour commenter le code, il faut utiliser « <!--> » pour du XML, « (:) » pour du XQuery et « /* */ » pour du JavaScript.

Dans le site, le visuel est géré par les fichiers JS, tandis que les données et le HTML sont gérés par .

III. La façon de créer un recueil et indexer des *exemples*

Créer un recueil

- Se rendre dans « Liste des recueils »
 - Cliquer sur « Créer un recueil »
 - Écrire le nom du recueil de préférence en langue originelle
 - Choisir l'édition scientifique sur laquelle se base l'indexation
 - Choisir les utilisateurs pouvant indexer les *exempla* du recueil

Créer un recueil

Titre:

Etablissement:

Langue: Date: Milieu: Zone géographique:

Informations

Nom d'utilisateur: Nom d'utilisateur: Nom d'utilisateur:

Historique

Nom d'utilisateur: Nom d'utilisateur: Nom d'utilisateur:

Actions

VALIDER

FIGURE B.1 – Créer un recueil

Modifier un recueil

- Se rendre dans la liste des recueils et en sélectionner un
 - Cliquer sur « Modifier »
 - Attendre la fin du chargement
 - Remplir les champs
 - Si les volets déroulant ne contiennent pas les informations nécessaires il faut se rendre dans « Admin des données » et les ajouter

The screenshot shows the 'Modifier le recueil' (Edit Collection) page for item TCo021. The interface is in French. The main fields include:

- Titre:** postulat
- Auteur:** André Véronneau
- Langue principale:** français
- Genre ou forme:** essai
- Zone géographique:** France
- Mots-clés:** (empty)
- Lien de traduction:** (empty)
- Notes:** (empty)
- Projet cette œuvre:** (empty)
- Sous-catégorie:** (empty)

At the top right are 'SÉPAREMENT' and 'SUPPRIMER' buttons.

FIGURE B.2 – Modifier un recueil

Créer ou modifier un *exemplum*

- Se rendre dans le tableau de bord ou dans le recueil créé
- Cliquer sur « Créeer un *exemplum* »
- Attendre la fin du chargement
- Compléter les champs
- Si les volets déroulant ne contiennent pas les informations nécessaires il faut se rendre dans « Admin des données » et les ajouter
- Enregistrer les modifications avec le bouton vert en haut à droite
- Pour supprimer un *exemplum* appuyer sur le bouton rouge « supprimer »

The screenshot shows the 'Modifier l'exemplum' (Edit exemplum) page for item TE000753. The interface is in French. The main fields include:

- Titre:** postulat
- Auteur:** André Véronneau, Biographe (THEMA)
- Image(s) (pdf, jpg):** (empty)
- Commentaire d'inscription:** (empty)
- Résumé:** (empty)
- Nom original:** (empty)
- Commentaire:** (empty)
- Sources:** (empty)

At the top right are 'SÉPAREMENT' and 'SUPPRIMER' buttons.

FIGURE B.3 – Créeer un *exemplum*

Annexe C

Code pour classer par ordre alphabétique les récits exemplaires avec un code Python

```
1  from pathlib import Path
2  from lxml import etree
3  import re
4
5  p = Path('../data')
6  exempla_variable = list(p.glob('TC*/*.xml'))
7  # print(exempla_variable)
8
9  def roman_to_arabic(roman):
10     """takes roman numbers (param roman) and returns arabic numbers"""
11     roman_numerals = {'I': 1, 'V': 5, 'X': 10, 'L': 50, 'C': 100, 'D': 500, 'M': 1000}
12     arabic = 0
13     prev_value = 0
14     for numeral in reversed(roman): # reversed affiche une variable dans le sens inverse.
15         value = roman_numerals[numeral]
16         if value < prev_value:
17             arabic -= value
18         else:
19             arabic += value
20         prev_value = value
21     return arabic
22
23  def lettre_en_numero(lettre):
24     """takes letter and returns its numeral position (a = 1, b = 2, etc)"""
25     lettre = lettre.lower()
26     if lettre.isalpha() and len(lettre) == 1:
27         return ord(lettre) - ord('a') + 1 # Passe par la table unicode. A partir de "a" (97) définit le numéro des autres lettres de l'alphabet
28     else:
29         return -1
30
31  def bis_to_two(texte):
32      """ takes "bis" and return "2" """
33  if texte.lower() == "bis":
34      return "2"
35
36
37
38  # Définir une liste de tuples contenant chaque regex et le traitement associé
39  regex_patterns_and_handlers = [
40
41      ### Sample : IV, 25 [82] == 82-1
42      (r"^\d+[.,\(\)] ?[\da-zA-z]+$", lambda match: match[0].replace(' ', '').replace('.', '-').replace(',', '-')),
43
44      ### Sample : 175b == 175-2
45      (r"^(\\d+)([A-Za-z])?$", lambda match: f"{match.group(1)}-{1 if not match.group(2) else f'{match.group(1)}-{lettre_en_numero(match.group(2))}'})",
46
47      #
48      (r"^\s*\d+(.)?\s*$", lambda match: match[0].replace(' ', '').replace('.','')),
49
50      ### Sample : XV, 15 == 15-15
51      (r"^(IVXLCDM)+ ?,[\s*(\\d+([\s*\d+)*)?,?]", lambda match: f'{roman_to_arabic(match[1].replace(' ', '-').replace(' ', ''))-{match[2].replace(' ', '-') if match[2] else ''}}'),
52
53      ### Sample : [52] == 52
54      (r"^\[(\\d+)([A-Za-z])?\]", lambda match: f'{match.group(1)}-{lettre_en_numero(match.group(2)) if match.group(2) else "1"}'),
```

```

56     ##### Sample : 1, 2, 3 == 1-2-3
57     (r"\d+(\.\d+)?(\.\s*\d+(\.\d+)?)*$", lambda match: re.sub(r'\s*,\s*', '-', match[0])),
58
59     ##### The Llanthony Stories : 25 == 25. The Llanthony Stories
60     ("The Llanthony Stories : (\d+$)", lambda match: match.group(1)),
61
62     ##### 720, 1-8 == 720. Ci nous dit
63     (r"(\d+)([A-Za-z]?)?", ?\d+-\d+", lambda match: f"{match.group(1)}-{lettre_en_numero(match.group(2)) if match.group(2) else '1'}"),
64
65     ##### Paris, BnF lat. 16481, Sermo 12, 2' == 12-2. Paris, BnF lat 16481-16482
66     ("Paris, BnF lat. (16481|16482), Sermo [A-Za-z]?(\d+)?:(, (\d+)?)", lambda match: f"{match.group(2)}-{match.group(3) if match.group(3) else ''}"),
67
68     ##### Sample : F13 == 1-13. Fabulae et parabolae
69     (r"(F|P|CS|CT)(\d+)([A-Za-z]?)?", lambda match: f"{1 if match.group(1) == 'F' else (2 if match.group(1) == 'P' else (3 if match.group(1) == 'CS' else 4))}-{match.group(2)}-1" if match.group(3) else f"{1 if match.group(1) == 'F' else (2 if match.group(1) == 'P' else (3 if match.group(1) == 'CS' else 4))}-{match.group(2)}"),
70
71     ##### Sample : I, 3 Exemplum 12 (Collet) == 12. Livre du jeu d'échecs
72     ("Exemplum ?(.)? (\d+).? .? ?\(\?C?c?ollet\)", lambda match: match.group(2)),
73
74     ##### Sample : n°1 == 1
75     (r"~n° ?(\d+$)", lambda match: match.group(1)), # "n°1"
76
77     # Sample : None = None
78     ("None$)", lambda match: match.group(1)), # Les cas de valeurs vide "None"
79
80     ##### Sample : RMEx_0126b = 0126-2. Érdy Codex
81     ("RMEx_(\d+)([a-zA-Z]?)?", lambda match: f"{match.group(1)}-{lettre_en_numero(match.group(2))}" if match.group(2) else match.group(1)), # "
82     RMEx_0126b
83
84     ##### Sample : 5 (1) == 5-1. Dits de Jehan de Saint-Quentin
85     (r"~(\d+)?\((\d+)\)\.?\$?", lambda match: f"{match.group(1)}-{match.group(2)}"), # 5 (1)
86
87     # Sample : Le dit de la bourjosse de Romme, p. 39-46 == 4.
88     (r"~([A-Za-z])\.?\.,? (L?D?)", lambda match: str(lettre_en_numero(match.group(1))),),
89
90     ##### Sample : Parabole 15B == 15-2. Parabolaire
91     (r"~Parabole (\d+)([A-Za-z]?)?", lambda match: f"{match.group(1)}-{lettre_en_numero(match.group(2))}" if match.group(2) else match.group(1)),
92
93     # Sample : exemple 49, p. 30 du texte hébreu == 49-1. Sefer Hamaassiyot
94     (r"~exemple ?(\d+)([A-Za-z]?)?(*.)?du texte hébreu?()?", lambda match: f"{match.group(1)}-{lettre_en_numero(match.group(2)) if match.group(2) else '1'}"),
95
96     ##### Sample : strophes 285-290 == 285. Libro de buen amor
97     (r"strophes (\d+)-(\d+)", lambda match: match.group(1)),
98
99     # Sample : XVI. Exemplum de tonellis olei == 16. Disciplina clericalis
100    (r"~([IVXLCDM]+)(.)? Exemplum de", lambda match: f"roman_to_arabic(match.group(1))"),
101
102    # Sample : fol. 4, n° 1 = 1. British Library, Add. 27909B
103    (r"~fol. (\d+)([A-Za-z]?)?,\.\.?-?(\d+)?([A-Za-z]?)?,\.\. (\d+)([A-Za-z]?)?", lambda match: f"{match.group(5)}-{lettre_en_numero(match.group(6)) if match.group(6) else '1'}" if match.group(6) else match.group(5)),
104
105    ##### Sample : OS013_E06 == 013-06. Sermones de sanctis Biga salutis intitulati
106    (r"~OS(\d+)_E(\d+$)", lambda match: f"{match.group(1)}-{match.group(2)}"),
107
108    # Sample : CPS-H 8 == 2-8
109    (r"~(?CP\$H\s)(\d+)([a-zA-Z]?),\s*pp?\.\.?s*(\d+)(?:-(\d+))?", lambda match: f"1-{match.group(1)}-{lettre_en_numero(match.group(2)) if match.group(2) else '1'}" if match.group(4) else f"{match.group(1)}" if not match.group(3) else f"{match.group(1)}-{match.group(3)}" if match.group(4) else f"{match.group(1)}" if not match.group(3) else f"{match.group(1)}-{lettre_en_numero(match.group(2)) if match.group(2) else ''}"), # Collectaneum
110
111    # Sample : 65 [b] == 65-2
112    (r"~(\d+)\s*(?:[\s*](\d+))?\$?", lambda match: f"{match.group(1)}-{lettre_en_numero(match.group(2))}", # 65 [b]
113
114    # Sample : Lettre 112, p. 276, 1. 9 - p. 277, 1. 8' == 112-276-1
115    (r"~Lettre (\d+), p\.\. ?(\d+), 1?L?\. (\d+)(?: -? -? p\.\. (\d+), 1\. (\d+))?", lambda match: f"{match.group(1)}-{match.group(2)}-{match.group(3)}"),
116
117    # Sample : Livre II, chapitre 28, col. 572 C == 2-28-572
118    (r"~Livre I{1,3}), chapitre (\d+), col. (\d+)(?:\s*[A-Z]\s*(?:-\s*[A-Z])?)?\$?", lambda match: f"roman_to_arabic(match.group(1))-{match.group(2)}-{match.group(3)}"),
119
120    # Sample : Livre I, chapitre 26, col. 537 B - 538 A == 1-26-537
121    (r"~Livre I{1,3}), chapitre (\d+), col. (\d+) [A-Za-z] - (\d+) [A-Za-z]", lambda match: f"roman_to_arabic(match.group(1))-{match.group(2)}-{match.group(3)}"),
122
123    # Sample : Predicazione della quaresima 1425 (Firenze S. Croce, 4 febbraio-8 aprile), XXXIX, 1. == 1425-39-1
124    (r"~.* (\d{4}) \(\?Firenze\?S\.\? Croce\, \d+ febbraio\-\d+ aprile\), ([IVXLCDM]+), (\d+)\.$", lambda match: f"{match.group(1)}-{roman_to_arabic(match.group(2))}-{match.group(3)}"),
125
126    # Predicazione della quaresima 1424 (Firenze, S. Croce, 8 marzo-3 maggio), XXIII, 2. == 1424-23-2
127    (r"~.* (\d{4}) \(\?Firenze\?S\.\? Croce\, \d+ marzo\-\d+ maggio\), ([IVXLCDM]+), (\d+)\.$", lambda match: f"{match.group(1)}-{roman_to_arabic(match.group(2))}-{match.group(3)}"),
128
129    # Sample : Prediche della primavera 1425 (Siena, chiesa di S. Francesco e Piazza del Campo, 20 aprile-10 giugno), VII, 6. == 1426-7-6
130    (r"~.* (\d{4}) \(\?Siena\? , chiesa di S\.\? Francesco e Piazza del Campo\, \d+ aprile\-\d+ giugno\), ([IVXLCDM]+), (\d+)\.$", lambda match: f"int(match.group(1))+1}-{roman_to_arabic(match.group(2))}-{match.group(3)}"),
131
132    # Sample : Bernardino da Siena, Prediche volgari sul Campo di Siena 1427 [ed. Delcorno, 1989], XXV, 5. == 1427-25-5
133    (r"~.* (\d{4}) \(\?ed\.\? Delcorno\, 1989\), ([IVXLCDM]+), (\d+)\.$", lambda match: f"{match.group(1)}-{roman_to_arabic(match.group(2))}-{match.group(3)}"),

```

```

133
134 # Sample : 'Lusurier repentant - exemplum ajouté après le chapitre II. 34 == 2-34
135 (r"(distinctio)?(chapitre)?(distinction)? ([IVXLCDM]+)\.?\s?(\d+)?\.$", lambda match: f"{roman_to_arabic(match.group(4))}-{match.group(5) if
136 match.group(5) else '1'}"),
137
138 # Sample : LES OEUVRES DE MISERICORDE, 890 == 890
139 (r"LES OEUVRES DE MISERICORDE, (\d+)", lambda match: match.group(1)),
140
141 # Sample : 273bis == 273-2
142 (r"(\d+) ?(bis)", lambda match: f"{match.group(1)}-{bis_to_two(match.group(2))}"),
143
144 # sample : Sermones De sanctis, [éd. Maggioni, en cours], 430a - 2 = 430-1-2
145 (r"Sermones De sanctis\, , [\éd. Maggioni, en cours]\,\, , (\d+)[A-Za-z] - (\d+)?", lambda match : f"{match.group(1)}-{lettre_en_numero(match.
146 group(2))}-{match.group(3)}"),
147
148 # Sample : Additiones: 6 == 6
149 (r"Additiones: (\d+)", lambda match: match.group(1)),
150
151 # Sample App., I, 1 == 1-1
152 (r"App.\, , ([IVXLCDM]+)\, , (\d+)", lambda match: f"{roman_to_arabic(match.group(1))}-{match.group(2)}"),
153
154 # Sample CPS-H 13 == 1-13
155 (r"~CPS-H (\d+)", lambda match: f"2-{match.group(1)}"),
156
157 # Sample 1, 21, 4b = 1-21-4-2
158 (r"~(\d+), (\d+), (\d+)([A-Za-z])", lambda match: f"{match.group(1)}-{match.group(2)}-{match.group(3)}-{lettre_en_numero(match.group(4))}"),
159
160 # Opusculum de naturis animalium, c. 9 (f. 8r) == 9-8
161 (r"Opusculum de naturis animalium, c\.\. (\d+) \,(f\.\. (\d+)[A-Za-z])", lambda match: f"{match.group(1)}-{match.group(2)}"),
162
163 # CORRIGER
164 (r"~p. ?(\d+)([A-Za-z])?([A-Za-z])? ~?([A-Za-z])?-? ~?([A-Za-z])?-(\d+)?([A-Za-z])?$', lambda match: '-'.join(filter(None, map(
165 lambda x: str(lettre_en_numero(x)) if x and x.isalpha() else x, match.groups())))), # Les sermons "p. 52b"
166
167 # Sample : Sermo 67 §19, p. 870 = 67-19-870. Sermons et visite pastorale
168 (r"~Sermo (\d+) § ?(\d+), p\.\. (\d+(?:-\d+)*$)", lambda match: f"{match.group(1)}-{match.group(2)}-{match.group(3).replace('-', '-')}" ),
169
170 file_namespace = {"tei": "http://www.tei-c.org/ns/1.0"}
171
172 # Pour calculer le pourcentage de modifications réalisées
173 total_titles = 0
174 unmodified_titles = 0
175
176 for file in exempla_variable:
177     tree = etree.parse(file)
178
179     title_index = tree.xpath(".///tei:teiHeader//tei:title/tei:desc[@type='title_index']", namespaces=file_namespace)
180
181     if len(title_index) > 0:
182         text_title = title_index[0].text
183         original_text_title = text_title # Stocker la valeur d'origine
184         modified = False # Variable de contrôle
185         for regex_pattern, handler in regex_patterns_and_handlers:
186             regex = re.compile(regex_pattern)
187             query = re.search(regex, str(text_title))
188             if query:
189                 text_title = handler(query)
190                 print(f"++++ '{original_text_title}' ('{query[0]}' = '{text_title}')")
191                 modified = True # Modification effectuée
192             break # Sortir de la boucle une fois une modification effectuée
193             if not modified: # Si aucune modification n'a été effectuée
194                 print(f"---- '{original_text_title}' (not modified)")
195             unmodified_titles += 1
196     total_titles += 1
197
198 percentage_unmodified = (unmodified_titles / total_titles) * 100
199 print(f"Percentage of unmodified titles: {percentage_unmodified}%")
200
201
202 # title_index[0].attrib['subtype'] = subtype_text
203 # permet de print la balise
204 # print(etree.tostring(title_index[0]))
205 # result_tree = etree.tostring(tree, encoding='utf-8', xml_declaration=False, doctype="", pretty_print=True)
206 # result_tree = result_tree.decode('utf-8')
207 # print(result_tree)
208 # with open(file, 'w') as writing:
209 #     writing.write(result_tree)
210 # print("Done")

```


Annexe D

Code pour classer par ordre alphabétique des récits exemplaires avec un code XQuery

1. Code pour vérifier la place des *exempla*

```
1      xquery version "3.1";
2
3      declare namespace tei="http://www.tei-c.org/ns/1.0";
4
5      let $output := 
6      let $collection-uri := "/db/apps/thema/data"
7
8      let $subcollections := xmldb:get-child-collections($collection-uri)
9
10     for $rec in $subcollections
11     let $rec-uri := $collection-uri || "/" || $rec
12     order by $rec
13     let $collection := collection($rec-uri)
14     let $exempla :=
15     for $exemplum in $collection/tei:TEI
16     order by $exemplum/@xml:id
17     return $exemplum
18     for $ex at $i in $exempla
19     (:  let $update := local:insert-number($ex/@xml:id, $i) :)
20     return
21     "UPDATED collection " || $rec-uri || " exempla ID " || $ex/@xml:id || " has position " || $i || " (" || $ex//tei:titleStmt/tei:
22     title/tei:desc[@type eq "title_index"] || ")"
23
24
25     return
xmldb:store("/db/apps/thema/data-test", "number_check.txt", string-join($output,"&#10;"))
```

2. Modification des données posant un problème

```
1      xquery version "3.1";
2
3      declare namespace tei="http://www.tei-c.org/ns/1.0";
4
5
6      let $collection-uri := "/db/apps/thema/data"
7      let $subcollections := xmldb:get-child-collections($collection-uri)
8      for $rec in $subcollections
9      let $rec-uri := $collection-uri || "/" || $rec
10     let $collection := collection($rec-uri)
11     let $exempla := $collection/tei:TEI
12     let $sorted-exempla :=
13     if ($rec = "TC0131") then
14     for $exemplum in $exempla
15     let $firstPart := fn:substring-before($exemplum, ',')
16     let $number := local:convert-to-number($firstPart)
17     order by $number ascending
18     return $exemplum
```

```

19     else if ($rec = "TC0155" or $rec = "TC0012" or $rec = "TC0013" or $rec = "TC0016" or $rec = "TC0105" or $rec = "TC0138" or
20             $rec = "TC0150" or $rec = "TC0014") then
21         for $exemplum in $exempla
22             let $num := local:extract-number(string($exemplum//tei:titleStmt/tei:title/tei:desc[@type eq "title_index"]))
23             order by $num ascending
24             return $exemplum
25         else if ($rec = "TC0142") then
26             for $exemplum in $exempla
27                 let $parts := tokenize($exemplum, '\s*')
28                 let $roman := $parts[1]
29                 let $arabic := local:roman-to-arabic($roman)
30                 let $subParts := subsequence($parts, 2)
31                 let $firstSubPart := if (count($subParts) > 0) then local:safe-convert-to-integer($subParts[1]) else ()
32                 let $secondSubPart := if (count($subParts) > 1) then local:safe-convert-to-integer($subParts[2]) else ()
33                 order by $arabic ascending, $firstSubPart ascending, $secondSubPart ascending
34                 return $exemplum
35             else
36                 for $exemplum in $exempla
37                     order by $exemplum/@xml:id
38                     return $exemplum
39                     for $ex at $i in $sorted-exempla
40                         let $update := local:insert-number($ex/@xml:id, $i)
41                     return $sorted-exempla
42
43
44
45
46
47     declare function local:insert-number($exempla-id as xs:string, $number as xs:integer ) {
48
49         let $exempla := collection("/db/apps/thema/data")/tei:TEI[@xml:id eq $exempla-id]
50         return
51         update insert attribute { "n" } { $number } into $exempla/tei:titleStmt/tei:title/tei:desc[@type eq "title_index"]
52
53     };
54
55     declare function local:convert-to-number($str as xs:string) as xs:string? {
56         if (not($str)) then ()
57         else
58             let $parts := tokenize($str, '\s*')
59             let $number := $parts[1]
60             let $letter := $parts[2]
61             let $letterValue := if ($letter) then xs:string(codepoints-to-string(string-to-codepoints($letter)) - 96) else ""
62             return if ($letterValue) then concat($number, ".", $letterValue) else $number
63     };
64
65     declare function local:extract-number($str as xs:string) as xs:integer? {
66         let $match := replace($str, ".*\((\d+)\)$", "$1")
67         return if ($match castable as xs:integer) then xs:integer($match) else ()
68     };
69
70     declare function local:roman-to-arabic($roman as xs:string) as xs:integer {
71         let $map := map {
72             'I': 1, 'IV': 4, 'V': 5, 'IX': 9, 'X': 10, 'XL': 40, 'L': 50, 'XC': 90, 'C': 100, 'CD': 400, 'D': 500, 'CM':
73             900, 'M': 1000
74         }
75         let $chars := fn:string-to-codepoints($roman)
76         let $values := for $char in $chars
77             return $map(fn:codepoints-to-string($char))
78             let $total := fn:fold-left($values, 0, function($acc, $current) {
79                 if ($acc >= $current) then $acc + $current else $current - $acc
80             })
81             return $total
82     };
83
84     declare function local:safe-convert-to-integer($str as xs:string) as xs:integer? {
85         let $numericPart := replace($str, "[^\d]", "") (: Supprime tout ce qui n'est pas un chiffre :)
86         return if ($numericPart castable as xs:integer) then xs:integer($numericPart) else ()
87     };

```

Annexe E

Fichiers XML de ThEMA et de SourcEncyMe

1. Exemple de fichier dans ThEMA

```
1 <?xml-model href=".../schema/tei_thema.rng" type="application/xml" schematypens="http://relaxng.org/ns/structure/1.0"?>
2 <TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:xi="http://www.w3.org/2001/XInclude" xmlns:dc="http://dublincore.org/documents/dcm -namespace/" xml:id="
3   TE002904" corresp="TC0011"
4     type="exemplum">
5       <teiHeader>
6         <fileDesc>
7           <titleStmt>
8             <title>
9               <desc type="custom_title"/>
10              <desc type="title_index">p. 27b</desc>
11            </title>
12          </titleStmt>
13          <publicationStmt>
14            <publisher>ThEMA (http://thema.huma-num.fr)</publisher>
15            <availability status="published">
16              <licence corresp="copyright-cc-by-nc-sa-4.0"/>
17            </availability>
18            <idno type="old_sql_id">2904</idno>
19          </publicationStmt>
20          <sourceDesc>
21            <list type="source_details">
22              <item type="source_text">
23                <p>
24                  ...
25                </p>
26              <item type="sources">
27                <p>
28                  ...
29                  <p>Feria quarta primae hebdomadae. Sermo I.</p>
30                </p>
31              <item type="commentary">
32                <p>
33                  ...
34                  <p>y</p>
35                </p>
36              <item type="allegory">y</item>
37            </list>
38            <listBibl>
39              <bibl type="manuscripts-editions" corresp="Z-IU7WR86C" n="t. I, p. 66-67"/>
40            </listBibl>
41            <list type="links">
42              <item type="link" corresp="http://sermones.net/thesaurus/document.php?id=jvor_210">Sermones.net</item>
43            </list>
44            <list type="linked_exempla"/>
45          </sourceDesc>
46        </fileDesc>
47        <profileDesc>
48          <textClass>
49            <keywords>
50              <term corresp="KW0039">KW0039</term>
51              <term corresp="KW0048">KW0048</term>
52              <term corresp="KW0155">KW0155</term>
53              <term corresp="KW0319">KW0319</term>
54              <term corresp="KW0435">KW0435</term>
55              <term corresp="KW0439">KW0439</term>
56              <term corresp="KW0458">KW0458</term>
57              <term corresp="KW0465">KW0465</term>
              <term corresp="KW0756">KW0756</term>
              <term corresp="KW0889">KW0889</term>
```

```

58             <term corresp="KW1118">KW1118</term>
59         </keywords>
60     </textClass>
61   </profileDesc>
62 <xenoData/>
63 <revisionDesc>
64   <listChange>
65     <change type="modify" resp="jrehr" when-custom="2019-08-06 14:14:00">eXist-db setup</change>
66     <change type="modify" resp="jrehr" when-custom="2020-11-20 20:45:55">transformation to new schema</change>
67   </listChange>
68 </revisionDesc>
69 </teiHeader>
70 <facsimile/>
71   <text>
72     <body>
73       <p xml:lang="de"/>
74       <p xml:lang="en"/>
75       <p xml:lang="es"/>
76       <p xml:lang="fr">Lodeur de la bonne réputation du Christ est assimilée aux odeurs du cèdre, de la myrrhe et de la
77           vigne qui font fuir les crapauds (représentation de la luxure), les serpents (représentation de 'lavarice) et les
78           vers (représentation de 'lorgueil). Les crapauds naissent souvent des organes génitaux des cadavres des hommes
79       .</p>
80     <p xml:lang="it"/>
81   </body>
82 </text>
83 </TEI>

```

2. Exemple de fichier dans SourcEncyMe

```

1 <TEI xmlns="http://www.tei-c.org/ns/1.0" xmlns:xxe="http://www.unicaen.fr/mrsh/pddn/xxe/1.0/" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
2   xmlns:xs="http://www.w3.org/2001/XMLSchema" xmlns:xi="http://www.w3.org/2001/XInclude" xmlns:ns="http://www.tei-c.org/ns/1.0" xmlns:hfp="http://
3   www.w3.org/2001/XMLSchema-hasFacetAndProperty" xmlns:examples="http://www.tei-c.org/ns/Examples" xmlns:dcr="http://www.isocat.org/ns/dcr" xmlns:
4   ad="http://ns.adobe.com/AdobeInDesign/5.0/" xmlns:a="http://ns.adobe.com/AdobeInDesign/4.0/" xsi:schemaLocation="http://www.tei-c.org/ns/1.0 http
5   ://www.unicaen.fr/mrsh/pddn/schemas/ichtya.xsd" xml:id="summa_de_exemplis_ac_similitudinibus_rerum_ed_anvers_1609">
6   <teiHeader>
7     <fileDesc>
8       <titleStmt>
9         <title ref="#liber_de_exemplis_et_similitudinibus_rerum_ioannis_de_sancto_geminiano">Summa de exemplis ac
10           similitudinibus rerum (éd. Anvers, 1609)</title>
11     <author>Iohannes de Sancto Geminiano</author>
12     <respStmt>
13       <persName id="Beatrice.Amelotti">Beatrice Amelotti</persName>
14       <resp>Mise en forme de l'édition</resp>
15     </respStmt>
16     <respStmt>
17       <persName id="Emmanuelle.Kuhry">Emmanuelle Kuhry</persName>
18       <resp>Transformation TEI</resp>
19     </respStmt>
20     <titleStmt>
21       <edition>Avertissement : Le texte du De exemplis de G. da S. Giminiano est mis provisoirement en ligne sur SourcEncyMe, car il est
22         régulièrement modifié et corrigé par parties ; il n'a pas encore bénéficié d'une relecture complète.</edition>
23     </titleStmt>
24     <publicationStmt>
25       <authority>François Bougard</authority>
26       <publisher>IRHT</publisher>
27       <pubPlace>Aubervilliers</pubPlace>
28       <date>2017</date>
29     </publicationStmt>
30     <sourceDesc>
31       <p/>
32     </sourceDesc>
33   </fileDesc>
34   <teiHeader>
35     <text>
36       <body>
37         <div type="oeuvre" xml:id="de_exemplis">
38           <div n="1" subtype="prologue" type="section1" xml:id="de_exemplis.prologue.1">
39             <head>DE EXEMPLIS ET RERUM SIMILITUDINIBUS LOCUPLETISSIMA. PROLOGUS.</head>
40             <cit xml:id="cit_idpi19503344">
41               <quote>Omnia facito secundum exemplar quod tibi monstratum est (inquit Apostolus). In omnibus operibus artium videmus, quod eorum opifices
42                 diriguntur et regulatur duplice exemplari: Nam qui sunt in arte periti, interius habent exemplar, scilicet formam artis, secundum
43                 quam operantur: qui vero artem addiscunt nondum eruditii in ipsa, aspiciunt ad exemplar exterius per aliquem artificem factum, sicut
44                 pueri qui scribere discunt, tenent pre oculis exemplar magistri.</quote>
45           </cit>
46         </body>
47       </div>
48     </text>
49   </teiHeader>
50 </TEI>

```

Annexe F

Code pour recherche dans les sources et les textes apparentés de ThEMA

Ce code ne représente pas la totalité du code présent dans le fichier search.xql. Pour économiser de la place je n'ai récupéré que le code qui permet de chercher dans les textes apparentés et dans les sources des *exempla*.

```
1  let $paramSces :=
2      if (request:get-parameter("sces", ()) != "") then
3          tokenize(request:get-parameter("sces", (), " "))
4      else ()
5
6
7
8      <div class="form-group row">
9          <label class="col-md-2 col-form-label col-form-label-sm" for="search-sources">{common:labels("search-sces")}</label>
10         <input type="text" class="col-md-8 form-control form-control-sm" id="search-sources" placeholder="exemplum sources search" name="sces"
11             accept-charset="UTF-8"/>
12     </div>
13
14
15     let $sources-in-exempla :=
16         if (count($paramSces) gt 0) (: Recherche seulement si $paramSces n'est pas vide :)
17             then
18                 let $query-sources :=
19                     <query>
20                         <bool>
21                             {
22                                 for $b in $paramSces
23                                     return <wildcard occur="must">{normalize-space(lower-case(sch:reduced-term($b)))}</wildcard>
24                             }
25                         </bool>
26                     </query>
27                     return ($ex-items//tei:sourceDesc/tei:list[@type="source_details"]/tei:item[@type="sources"]/tei:p[ft:query(.,$query-
28                         sources,$ftqueryparams)])
29                 else (: Pas de recherche si $paramSces est vide :)
30                 $ex-items
31             return
32
33
34     (<div>
35         <div class="d-flex justify-content-between border-bottom align-items-center mb-3">
36             <h2 class="font-cormorant-garamond">{common:labels("search-results")}</h2>
37         {
38             if (count($final-ex-items) gt 0)
39                 then concat(" ",count($final-ex-items), " exempla")
40             else (),
41             if (count($sources-in-exempla) gt 0)
42                 then concat(" ",count($sources-in-exempla), " sources")
43             else ()
44         }
45         </h2>
46         <a class="btn btn-primary" href="{{$g:URLsearch}}>{common:labels("new-search")}</a>
47     </div>
48
49
50
```

```

51 <div class="row md-12 mb-2 d-flex align-items-center justify-content-between alert {if (count($final-ex-items) = 0 or count($sources-in-exempla
52     ) = 0)
53         then "alert-danger"
54         else "alert-success"}>
55
56
57 if (count($paramSces) gt 0)
58     then
59         (<strong>{common:labels("sourcessources")}</strong>, concat(" = ", string-join($paramSces, " "), " "))
60     else(),
61
62
63
64 if (count($final-ex-items) = 0 or count($sources-in-exempla) = 0)
65     then
66         (<br/>,<strong>{common:labels("search-no-results")}</strong>)
67     else ()
68
69
70
71 if ((count($final-ex-items) gt 0) or (count($sources-in-exempla) = 0))
72     then
73         (<a type="button" class="btn btn-primary" href="{concat($g:URLget,"?objecttype=searchresult&objectformat=csv&request:
74             get-query-string())}" target="_blank">{common:labels("download")}.csv}</a>
75     else ()
76
77
78 if (count($sources-in-exempla) gt 0)
79     then
80         <div class="row">
81             <div class="col-sm-12">
82                 <table id="tableSearch_in_sources" class="table table-sm responsive dataTable no-footer dtr-inline" style="width: 100%; data-page-length="50" role="grid" aria-describedby="myTable_info">
83                     <thead>
84                         <tr>
85                             <th>{common:labels("id-coll")}</th>
86                             <th>{common:labels("id-ex")}</th>
87                             <th>{common:labels("author")}</th>
88                             <th>{common:labels("title")}</th>
89                             <th>{common:labels("exempla")}</th>
90                             <th>{common:labels("keywords")}</th>
91                     </tr>
92                 </thead>
93                 <tbody>
94                     {
95                         for $sources-exempla in $sources-in-exempla
96                             let $fullsources := if (count($paramSces) gt 0) then $sources-exempla/ancestor::tei:TEI else
97                                 $sources-exempla
98                             let $coll := $coll-items/id($fullsources/@corresp)
99                             return
100                             <tr>
101                                 <td><a href="${$g:URLcollections || $fullsources/data(@corresp)}">${$fullsources/
102                                     data(@corresp)}${$fullsources/data(
104                                         @xml:id)}

```

Annexe G

Code pour comparer les textes de ThEMA et de SourcEncyMe

```
1 import pandas as pd
2 from tqdm import tqdm
3 from sentence_transformers import SentenceTransformer, util
4
5 # Load the SBERT model for French
6 model = SentenceTransformer('sentence-transformers/paraphrase-multilingual-MiniLM-L12-v2')
7
8 # Function to compute similarity between two embeddings
9 def compute_similarity(embedding1, embedding2):
10    try:
11        # Compute cosine similarity
12        similarity = util.pytorch_cos_sim(embedding1, embedding2).item()
13        return similarity
14    except Exception as e:
15        print(f"An error occurred while computing similarity: {e}")
16        return None
17
18 # Your manually entered text
19 text1 = "De même, l'amour de la charité transforme même le mal en bien. En effet, comme le dit l'Apôtre : 'Pour ceux qui aiment Dieu, tout contribue au bien.' Un exemple en est le poirier, car les poires immatures, qui sont acides, dures et grossières, et mauvaises à manger, deviennent bonnes après cuisson au feu. De la même manière, par la cuisson et la vigueur du feu de la charité, même les grandes et lourdes adversités se transforment en bien pour l'homme, du moins pour l'âme, grâce à la patience que la charité engendre. En effet, la charité est patiente, comme le dit l'Apôtre."
20
21 # Try to read the Excel file
22 try:
23     df = pd.read_excel('french_texts_.xlsx')
24 except Exception as e:
25     print(f"An error occurred while reading the Excel file: {e}")
26     exit()
27
28 # Ensure that the DataFrame has the expected structure
29 if len(df.columns) < 2:
30     print("The DataFrame should have at least two columns.")
31     exit()
32
33 # Get the texts from the specified column
34 texts_from_excel = df.iloc[:, 1].dropna().astype(str) # Assuming the texts are in the second column
35
36 # Compute embedding for the manually entered text
37 embedding1 = model.encode(text1, convert_to_tensor=True)
38
39 # Compute embeddings for the texts from the Excel file in batches
40 batch_size = 32
41 similarities = []
42 for i in tqdm(range(0, len(texts_from_excel), batch_size), desc="Calculating Similarities"):
43     batch_texts = texts_from_excel[i:i+batch_size].tolist()
44     batch_embeddings = model.encode(batch_texts, convert_to_tensor=True)
45
46 # Compute similarities for each text in the batch
47 for text2, embedding2 in zip(batch_texts, batch_embeddings):
48     similarity = compute_similarity(embedding1, embedding2)
49     if similarity is not None:
50         similarities.append((text2, similarity))
51
52 # Check if similarities were calculated
53 if not similarities:
54     print("No similarities were calculated. Exiting...")
55     exit()
56
57 # Sort the similarities by similarity score in descending order
```

```
58     similarities.sort(key=lambda x: x[1], reverse=True)
59
60     # Create a DataFrame from the sorted similarities
61     result_df = pd.DataFrame(similarities, columns=['Text', 'Similarity'])
62
63     # Save the DataFrame to an Excel file sorted by similarity score
64     try:
65         result_df.to_excel('comparaison2vite.xlsx', index=False)
66         print("Results saved successfully!")
67     except Exception as e:
68         print(f"An error occurred while saving the results to Excel: {e}")
```

Annexe H

Feuille de style XSL pour repérer les *exempla* de SourcEncyMe

```
1  <xsl:stylesheet version="1" xmlns:xsl="http://www.w3.org/1999/XSL/Transform" xmlns:tei="http://www.tei-c.org/ns/1.0" exclude-result-prefixes="tei">
2      <xsl:output method="xml" indent="yes" omit-xml-declaration="yes" encoding="UTF-8"/>
3
4
5      <xsl:template match="node()|@*">
6          <xsl:copy>
7              <xsl:apply-templates select="node()|@*"/>
8          </xsl:copy>
9      </xsl:template>
10
11
12      <xsl:template match="tei:cit">
13          <xsl:copy>
14              <xsl:apply-templates select="@*"/>
15              <xsl:apply-templates select="tei:bibl"/>
16              <xsl:if test="tei:quote[matches(., 'exemplum|Exemplum')] and not(tei:bibl[tei:ref[@target='#exemplum'])]">
17                  <xsl:element name="bibl">
18                      <xsl:element name="ref" >
19                          <xsl:attribute name="target">#exemplum</xsl:attribute>
20                          <xsl:attribute name="cert">low</xsl:attribute>
21                          <xsl:text>Exemplum</xsl:text>
22                  </xsl:element>
23              </xsl:element>
24          </xsl:if>
25          <xsl:apply-templates select="tei:quote"/>
26      </xsl:copy>
27  </xsl:template>
28
29
30      <xsl:template match="tei:quote">
31          <xsl:copy>
32              <xsl:apply-templates select="@*|node()"/>
33          </xsl:copy>
34  </xsl:template>
35
36
37      <xsl:template match="tei:quote/text()">
38          <xsl:variable name="text" select=".//"/>
39          <xsl:for-each select="tokenize($text, '\s+')">
40              <xsl:choose>
41                  <xsl:when test="matches(., 'exemplum', 'i')">
42                      <seg type="marqueur">
43                          <xsl:value-of select=".//"/>
44                      </seg>
45                  </xsl:when>
46                  <xsl:otherwise>
47                      <xsl:value-of select=".//"/>
48                  </xsl:otherwise>
49              </xsl:choose>
50              <xsl:if test="position() != last()">
51                  <xsl:text> </xsl:text>
52              </xsl:if>
53          </xsl:for-each>
54  </xsl:template>
55
56      <xsl:template name="wrap-exemplum">
57          <xsl:param name="text"/>
58          <xsl:choose>
59              <xsl:when test="contains($text, 'exemplum') or contains($text, 'Exemplum')">
60                  <xsl:variable name="before" select="substring-before($text, 'exemplum')"/>
```

```
61          <xsl:variable name="after" select="substring-after($text, 'exemplum')"/>
62          <xsl:value-of select="$before"/>
63          <seg type="marqueur">
64          <xsl:text>exemplum</xsl:text>
65          </seg>
66          <xsl:call-template name="wrap-exemplum">
67              <xsl:with-param name="text" select="$after"/>
68          </xsl:call-template>
69      </xsl:when>
70      <xsl:otherwise>
71          <xsl:value-of select="$text"/>
72      </xsl:otherwise>
73  </xsl:choose>
74 </xsl:template>
75
76 </xsl:stylesheet>
```

Annexe I

Code pour indexer automatiquement les *exempla* dans ThEMA

```
1 import xml.etree.ElementTree as ET
2 from selenium import webdriver
3 from selenium.webdriver.common.by import By
4 from selenium.webdriver.firefox.service import Service as FirefoxService
5 from selenium.webdriver.support.ui import WebDriverWait
6 from selenium.webdriver.support import expected_conditions as EC
7 import time
8 from bs4 import BeautifulSoup
9
10 # Chemin vers le geckodriver sur votre système
11 geckodriver_path = '/snap/bin/geckodriver' # Mettez à jour ce chemin si nécessaire
12
13 # Initialiser le service Firefox avec le chemin vers geckodriver
14 service = FirefoxService(executable_path=geckodriver_path)
15
16 # Initialiser le pilote WebDriver pour Firefox
17 driver = webdriver.Firefox(service=service)
18
19 def extract_quotes_from_xml(xml_file):
20     with open(xml_file, 'r', encoding='utf-8') as file:
21         content = file.read()
22
23     soup = BeautifulSoup(content, 'xml')
24
25     quotes = []
26     for cit in soup.find_all('cit'):
27         ref = cit.find('ref', target="#exemplum")
28         if ref is not None and ref.text == 'Exemplum':
29             quote = cit.find('quote')
30             if quote is not None:
31                 quote_text = ''.join(quote.stripped_strings)
32                 if quote_text:
33                     quotes.append(quote_text)
34
35     return quotes
36
37 # Exemple d'utilisation
38 xml_file_path = 'speculum_historiale_31_07_2024.xml' # Mettez à jour ce chemin si nécessaire
39 quotes = extract_quotes_from_xml(xml_file_path)
40
41 try:
42     # URL de la page de connexion
43     login_url = "http://127.0.1.1:8080/exist/apps/thema/login"
44     driver.get(login_url)
45
46     # Attendre que la page se charge
47     WebDriverWait(driver, 20).until(EC.presence_of_element_located((By.ID, 'user')))
48
49     # Remplir le nom d'utilisateur
50     username_input = driver.find_element(By.ID, 'user')
51     username_input.send_keys("admin")
52
53     # Remplir le mot de passe
54     password_input = driver.find_element(By.ID, 'password')
55     password_input.send_keys("admin")
56
57     # Soumettre le formulaire en utilisant .submit() sur le formulaire lui-même
58     login_form = driver.find_element(By.ID, 'form')
59     login_form.submit()
60
61     # Attendre que la page se charge après la connexion
```

```

62 WebDriverWait(driver, 20).until(EC.url_changes(login_url))
63
64 # Utiliser les cookies pour rester connecté lors de futures visites
65 cookies = driver.get_cookies()
66 for cookie in cookies:
67     driver.add_cookie(cookie)
68
69 # Naviguer vers l'onglet "Collections"
70 collections_url = "http://127.0.1.1:8080/exist/apps/thema/collections/TC0180"
71 driver.get(collections_url)
72
73 # Attendre que la page TC0179 se charge
74 WebDriverWait(driver, 20).until(EC.presence_of_element_located((By.XPATH, '//a[@id="profile-tab"]')))
75
76 # Trouver le bouton "Liste des exempla" et cliquer dessus
77 liste_exempla_button = driver.find_element(By.XPATH, '//a[@id="profile-tab"]')
78 liste_exempla_button.click()
79
80 # Aller à la page de création d'exempla
81 creat_exempla_url = "http://127.0.1.1:8080/exist/apps/thema/edit?exemplum=new&addcoll=TC0180"
82
83 for index, quote in enumerate(quotes, start=1):
84     driver.get(creat_exempla_url)
85
86     # Attendre que la page se charge
87     WebDriverWait(driver, 20).until(EC.presence_of_element_located((By.ID, 'source_text')))
88
89     # Remplir le champ source_text avec la citation extraite
90     source_text_input = driver.find_element(By.ID, 'source_text')
91     source_text_input.clear()
92     source_text_input.send_keys(quote)
93
94     # Remplir le champ title_for_exempla avec le numéro d'exemplum (index)
95     title_for_exempla_input = driver.find_element(By.ID, 'title_for_exempla')
96     title_for_exempla_input.clear()
97     title_for_exempla_input.send_keys(str(index))
98
99     # Cliquer sur le bouton "btn-save-collection"
100    save_button = driver.find_element(By.ID, 'btn-save-collection')
101    save_button.click()
102
103    # Attendre un peu pour que les actions se terminent et le formulaire se réinitialise
104    time.sleep(5)
105
106    # Retourner à la page d'accueil du recueil
107    collec_url = "http://127.0.1.1:8080/exist/apps/thema/collections/TC0180"
108    driver.get(collec_url)
109
110    list_exe_button = driver.find_element(By.XPATH, '//a[@id="profile-tab"]')
111    list_exe_button.click()
112
113 finally:
114     # Fermer le navigateur à la fin du script
115     driver.quit()
116

```

Annexe J

Code Python pour ajouter les liens dans les XML des deux bases de données

```
1 import os
2 import logging
3 from bs4 import BeautifulSoup
4 from difflib import SequenceMatcher
5
6 # Configuration du logger
7 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
8
9 def extract_quotes_from_xml(xml_file):
10     try:
11         with open(xml_file, 'r', encoding='utf-8') as file:
12             content = file.read()
13     except Exception as e:
14         logging.error(f"Erreur lors de la lecture du fichier {xml_file}: {e}")
15         return []
16
17     soup = BeautifulSoup(content, 'xml')
18
19     quotes = []
20     for cit in soup.find_all('cit'):
21         cit_id = cit.get('xml:id')
22         ref = cit.find('ref', target="#exemplum")
23         if ref and ref.text.strip() == 'Exemplum':
24             quote = cit.find('quote')
25             if quote:
26                 quote_text = ''.join(quote.stripped_strings).strip()
27                 if quote_text:
28                     quotes.append((cit_id, quote_text))
29
30     return quotes
31
32 def similarity(text1, text2):
33     return SequenceMatcher(None, text1, text2).ratio()
34
35 def ajouter_liens_dans_xml_origine(xml_origine_file, dossier_tc0179_origine):
36     quotes = extract_quotes_from_xml(xml_origine_file)
37
38     try:
39         with open(xml_origine_file, 'r', encoding='utf-8') as file:
40             content = file.read()
41     except Exception as e:
42         logging.error(f"Erreur lors de la lecture du fichier {xml_origine_file}: {e}")
43         return
44
45     soup_origine = BeautifulSoup(content, 'xml')
46     tei_ns = {'tei': 'http://www.tei-c.org/ns/1.0'}
47
48     for quote_id, quote_text in quotes:
49         for filename in os.listdir(dossier_tc0179_origine):
50             if filename.endswith('.xml'):
51                 xml_file_path = os.path.join(dossier_tc0179_origine, filename)
52                 try:
53                     with open(xml_file_path, 'r', encoding='utf-8') as file:
54                         content_tc0179 = file.read()
55                 except Exception as e:
56                     logging.error(f"Erreur lors de la lecture du fichier {xml_file_path}: {e}")
```

```

58         continue
59
60     soup_tc0179 = BeautifulSoup(content_tc0179, 'xml')
61     source_text_item = soup_tc0179.find('item', {'type': 'source_text'})
62     if source_text_item:
63         source_text_p = source_text_item.find('p', {'xml:lang': ''})
64         if source_text_p:
65             source_text = ''.join(source_text_p.stripped_strings).strip()
66             sim_score = similarity(quote_text, source_text)
67             if sim_score > 0.8:
68                 logging.info(f"Ajout de lien dans {xml_origine_file} pour la citation avec l'id {quote_id} vers
69                             le fichier {filename}.")
70                 cit = soup_origine.find('cit', {'xml:id': quote_id})
71                 if cit:
72                     ref = cit.find('ref', target="#exemplum", cert="low")
73                     if ref:
74                         link = f"https://thema.huma-num.fr/exempla/{os.path.splitext(filename)[0]}"
75                         if 'corresp' in ref.attrs:
76                             ref['corresp'] += f' {link}'
77                         else:
78                             ref['corresp'] = link
79
80     # Suppression des préfixes de namespace indésirables
81     xml_str = str(soup_origine)
82     xml_str = xml_str.replace('<ns:', '<').replace('</ns:', '</').replace('xmlns:ns="http://www.tei-c.org/ns/1.0"', '')
83
84     try:
85         with open(xml_origine_file, 'w', encoding='utf-8') as file:
86             file.write(xml_str)
87         logging.info(f"Fichier {xml_origine_file} mis à jour avec les nouveaux liens.")
88     except Exception as e:
89         logging.error(f"Erreur lors de l'écriture du fichier {xml_origine_file}: {e}")
90
91 def modifier_fichiers_xml(xml_origine_file, dossier_tc0179_origine, dossier_tc0179_modifie):
92     quotes = extract_quotes_from_xml(xml_origine_file)
93
94     if not os.path.exists(dossier_tc0179_modifie):
95         os.makedirs(dossier_tc0179_modifie)
96
97     for filename in os.listdir(dossier_tc0179_origine):
98         if filename.endswith('.xml'):
99             xml_file_path = os.path.join(dossier_tc0179_origine, filename)
100            with open(xml_file_path, 'r', encoding='utf-8') as file:
101                content = file.read()
102
103                soup = BeautifulSoup(content, 'xml')
104                modified = False
105
106                for quote_id, quote_text in quotes:
107                    source_text_item = soup.find('item', {'type': 'source_text'})
108                    if source_text_item:
109                        source_text_p = source_text_item.find('p', {'xml:lang': ''})
110                        if source_text_p:
111                            source_text = ''.join(source_text_p.stripped_strings).strip()
112
113                            # Calcul de la similarité entre le texte de la citation et le texte source du fichier XML
114                            sim_score = similarity(quote_text, source_text)
115                            if sim_score > 0.8: # Ajustez ce seuil selon vos besoins
116                                print(f"Modification trouvée dans {filename} pour le texte de quote avec l'id {quote_id}.")
117                                links_list = soup.find('list', {'type': 'links'})
118                                if not links_list:
119                                    source_desc = soup.find('sourceDesc')
120                                    if source_desc:
121                                        links_list = soup.new_tag('list', type='links')
122                                        source_desc.append(links_list)
123                                        links_list.append('\n')
124
125                                    new_item = soup.new_tag('item', type='link', corresp=f"http://sourcencyme.irht.cnrs.fr/
126                                            encyclopedie/speculum_historiale_version_sm_trifaria_ms_douai_bm_797?citid={quote_id}")
127                                    new_item.string = 'SourcEncyMe'
128                                    links_list.append(new_item)
129                                    links_list.append('\n')
130                                    modified = True
131
132                                new_xml_path = os.path.join(dossier_tc0179_modifie, filename)
133                                if modified:
134                                    # Suppression des préfixes de namespace indésirables
135                                    xml_str = str(soup)
136                                    xml_str = xml_str.replace('<ns:', '<').replace('</ns:', '</').replace('xmlns:ns="http://www.tei-c.org/ns/1.0"', '')
137
138                                    with open(new_xml_path, 'w', encoding='utf-8') as file:
139                                        file.write(xml_str)
140                                        print(f"Fichier modifié écrit : {new_xml_path}")
141
142    # Chemins des fichiers et dossiers
143    xml_origine_file = 'speculum_historiale_31_07_2024.xml'
144    dossier_tc0179_origine = 'TC0180'
145    dossier_tc0179_modifie = 'TC0180_modif'
146
147    # Appel de la fonction pour ajouter des liens dans le fichier XML d'origine

```

```
149 ajouter_liens_dans_xml_origine(xml_origine_file, dossier_tc0179_origine)
150 # Appel de la fonction pour modifier les fichiers XML
152 modifier_fichiers_xml(xml_origine_file, dossier_tc0179_origine, dossier_tc0179_modifie)
```


Annexe K

Code Python pour créer des liens entre les deux bases en cas de mention d'encyclopédies dans ThEMA

```
1 import os
2 import xml.etree.ElementTree as ET
3 import pandas as pd
4 import re
5
6 # Définir l'espace de noms TEI
7 TEI_NAMESPACE = 'http://www.tei-c.org/ns/1.0'
8 NS_MAP = {'tei': TEI_NAMESPACE}
9 ET.register_namespace('', TEI_NAMESPACE) # Enregistrer l'espace de noms par défaut
10
11 def roman_to_arabic(roman):
12     """
13     Convertir les chiffres romains en chiffres arabes.
14     """
15     roman_numerals = {'I': 1, 'V': 5, 'X': 10, 'L': 50, 'C': 100, 'D': 500, 'M': 1000}
16     arabic = 0
17     prev_value = 0
18     for char in reversed(roman):
19         value = roman_numerals.get(char, 0)
20         if value < prev_value:
21             arabic -= value
22         else:
23             arabic += value
24         prev_value = value
25     return arabic
26
27 def extract_numbers_from_text(text, search_term):
28     """
29     Extraire les numéros à partir du texte en utilisant l'expression régulière spécifiée.
30     Convertir les chiffres romains en chiffres arabes et ajouter 1 au premier numéro.
31     """
32     numbers = {'numero1': '', 'numero2': ''}
33     match = re.search(search_term, text)
34     if match:
35         num1 = match.group(1) if match.group(1) else ''
36         num2 = match.group(2) if match.group(2) else ''
37
38         # Convertir le premier numéro en chiffres arabes s'il est en chiffres romains
39         if num1 and re.match(r'^[IVXLCDM]+$', num1):
40             num1_arabic = roman_to_arabic(num1)
41             numbers['numero1'] = str(num1_arabic + 1) # Ajouter 1 à num1
42         elif num1.isdigit():
43             numbers['numero1'] = str(int(num1) + 1) # Ajouter 1 à num1 s'il est déjà en chiffres arabes
44         else:
45             numbers['numero1'] = num1
46
47         numbers['numero2'] = num2
48     return numbers
49
50 def check_numbers_in_other_file(numbers, other_xml_root, base_xml_id):
51     """
52     Vérifier si les numéros extraits existent dans le fichier XML autre.
```



```
145     'file_path': file_path,
146     'paragraph': item.text.strip(),
147     'numero1': numbers['numero1'],
148     'numero2': numbers['numero2'],
149     'cit_id': cit_id,
150     'base_xml_id': base_xml_id
151   })
152   print(f"Correspondance trouvée :\nFichier : {file_path}\nParagraphe : {item.text.strip()
153   ()}\nNuméro 1 : {numbers['numero1']}\nNuméro 2 : {numbers['numero2']}\nCIT ID : {
154   cit_id}\nBase XML ID : {base_xml_id}\n")
155   # Ajouter le cit_id au fichier
156   add_cit_id_to_file(file_path, cit_id)
157
158   except (ET.ParseError, Exception) as e:
159     print(f"Erreur avec le fichier {file_path}: {e}")
160
161   # Enregistrer le fichier modifié other_file_path
162   other_tree.write(other_file_path, encoding='utf-8', xml_declaration=True)
163
164   return matching_paragraphs
165
166   # Exemple d'utilisation :
167   directory_path = 'data' # Assurez-vous que ce chemin est correct
168   search_term = r'S?[^peculum historiale,? ([IVXLCDM]+|\d+)[.,]?' # Terme à rechercher avec regex
169   other_file_path = 'vincentius_belvicensis-speculum_historiale_version_sm_trifaria_ms_douai_bm_797.xml' # Chemin vers l'autre fichier XML TEI à
170   # vérifier
171
172   matching_paragraphs = search_in_tei_files(directory_path, search_term, other_file_path)
173
174   # Créer un DataFrame pandas pour les résultats
175   df = pd.DataFrame(matching_paragraphs)
176
177   # Exporter les résultats dans un fichier Excel
178   output_file = 'resultats_recherche.xlsx'
179   df.to_excel(output_file, index=False)
180
181   print(f"Les résultats de la recherche ont été exportés vers {output_file}.")
```


Index

API, 2, 43

CRH, vi, 30

EHESS, ix

Emmanuelle Kuhry, vii, 15, 21, 26, 28

Encyclopédies, i, vi, vii, 1, 2, 6, 9–11, 15–21,
23, 25, 26, 28, 31–33

IRHT, vii, ix

Isabelle Draelants, vii, ix, 21, 23, 28, 29

LabEx HaStec, v, ix

Marie-Anne Polo de Beaulieu, vi, ix, 5, 23,
29, 30

Mementos, 7–9

Open Data, i, v, 29, 35

Python, i, 6, 24–26, 47

Récits exemplaires, i, vi, viii, 1, 2, 6, 7, 13,
15–17, 20, 21, 23, 24, 29–33, 51

SGBD, viii, ix, 1

Speculum historiale, 26, 30

Speculum naturale, 26

Vincent de Beauvais, vi, viii, 30

XML, 2, 5, 6, 11, 21, 24–26, 30, 35, 43, 53,
54

XML-TEI, i, vi–ix, 1, 2, 35

XQuery, i, viii, 6, 24, 44, 45

XSL, 21, 30

Table des matières

Résumé	i
Remerciements	iii
Introduction	ix
I. Présentation des bases de données ThEMA et SourcEncyMe	ix
II. Motivations et enjeux de l’interconnexion des bases de données	xi
III. Présentation de la problématique et du plan	xiii
1 Réflexions sur les différentes possibilités d’interconnexion	1
I. Solution radicale : fusionner les deux bases de données	1
II. Solution idéale : développer une API	2
III. Un compromis : l’ajout de liens simples	2
2 Stratégies pour l’intégration des liens dans les deux bases de données	5
I. Vérification de la qualité des données et de leur pérennité	5
II. Identifications des liens entre les bases de données	6
III. Intégration des liens XML dans les bases de données	9
1. Recherche des emplacements et des informations dans les XML pour la création des liens dans ThEMA	9
2. Placement des liens et des visualisation des <i>exempla</i> dans les XML de SourcEncyMe	11
3 Repérages et recherches des données pour réaliser la connexion	15
I. Création d’outils pour chercher dans les données qui seront connectées . .	15
II. Identification des liens par comparaison des textes	17
III. Identification des <i>exempla</i> via des mots-clés	20
4 Crédit des liens dans les bases de données	23
I. Indexation des <i>exempla</i> de SourcEncyMe dans ThEMA	23
II. Gestion des liens vers SourcEncyMe dans ThEMA	24
1. Création des liens	24

2.	Perspectives pour la création du lien	25
III.	Création de liens vers SourcEncyMe depuis les mentions d'encyclopedies dans ThEMA	26
5	L'interconnexion des bases de données comme moteur de l'Open Data dans les sciences	29
I.	Renforcement des collaborations scientifiques	29
II.	Valorisation des bases de données	30
III.	Amélioration de l'expérience utilisateur	30
6	De nouvelles perspectives grâce à l'interconnexion	33
I.	Bénéfices pour la recherche de la mise en relation des deux bases de données	33
II.	Techniques de visualisation pour les relations entre récits exemplaires et encyclopédies	34
Conclusion		38
A	Installation de ThEMA	39
B	Fonctionnement de ThEMA	43
I.	La structure de ThEMA	43
II.	Le fonctionnement de ThEMA	44
III.	La façon de créer un recueil et indexer des <i>exempla</i>	45
C	Code pour classer par ordre alphabétique des récits exemplaires avec un code Python	47
D	Code pour classer par ordre alphabétique des récits exemplaires avec un code XQuery	51
1.	Code pour vérifier la place des <i>exempla</i>	51
2.	Modification des données posant un problème	51
E	Fichiers XML de ThEMA et de SourcEncyMe	53
1.	Exemple de fichier dans ThEMA	53
2.	Exemple de fichier dans SourcEncyMe	54
F	Code pour recherche dans les sources et les textes apparentés de ThEMA	55
G	Code pour comparer les textes de ThEMA et de SourcEncyMe	57
H	Feuille de style XSL pour repérer les <i>exempla</i> de SourcEncyMe	59
I	Code pour indexer automatiquement les <i>exempla</i> dans ThEMA	61

J	Code Python pour ajouter les liens dans les XML des deux bases de données	63
K	Code Python pour créer des liens entre les deux bases en cas de mention d'encyclopédies dans ThEMA	67