



HAL
open science

Thesaurus Exemplorum Medii Aevi: une base de données collaborative sur les exempla médiévaux

Jean-Paul Rehr, Marie-Anne Polo de Beaulieu

► To cite this version:

Jean-Paul Rehr, Marie-Anne Polo de Beaulieu. Thesaurus Exemplorum Medii Aevi: une base de données collaborative sur les exempla médiévaux. *Humanités numériques*, 2021, *Varia*, 2021-4 (4), 10.4000/revuehn.2630 . hal-03500584

HAL Id: hal-03500584

<https://hal.science/hal-03500584v1>

Submitted on 22 Dec 2021

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

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



Distributed under a Creative Commons Attribution 4.0 International License

Thesaurus Exemplorum Medii Aevi : une base de données collaborative sur les *exempla* médiévaux

Thesaurus Exemplorum Medii Aevi: A Collaborative Database of Medieval Exempla

Jean-Paul Rehr et Marie-Anne Polo de Beaulieu



Édition électronique

URL : <https://journals.openedition.org/revuehn/2630>

ISSN : 2736-2337

Éditeur

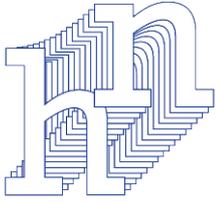
Humanistica

Référence électronique

Jean-Paul Rehr et Marie-Anne Polo de Beaulieu, « *Thesaurus Exemplorum Medii Aevi* : une base de données collaborative sur les *exempla* médiévaux », *Humanités numériques* [En ligne], 4 | 2021, mis en ligne le 01 décembre 2021, consulté le 21 décembre 2021. URL : <http://journals.openedition.org/revuehn/2630>



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la Licence Creative Commons Attribution 4.0 International.



Thesaurus Exemplorum Medii Aevi : une base de données collaborative sur les exempla médiévaux

Thesaurus Exemplorum Medii Aevi: A Collaborative Database of Medieval Exempla

Jean-Paul Rehr et Marie-Anne Polo de Beaulieu

Résumés

Les sermons médiévaux renferment des milliers d'*exempla*, ou histoires exemplaires, souvent utilisés pour transmettre doctrine et morale religieuses aux fidèles. Ils ont longtemps été considérés comme une part négligeable des textes dogmatiques et moraux, et l'attention s'est focalisée sur eux en tant que sous-ensemble des contes populaires et des fables dans les études littéraires. Avec le tournant anthropologique, une nouvelle approche des *exempla* en tant que sources culturelles et historiques a été développée, notamment par l'historien Jacques Le Goff et ses héritiers.

ThEMA, une base de données des *exempla* médiévaux, a commencé sa vie dans les années 1990 entre les mains de ces héritiers, et elle s'est développée et transformée jusqu'à nos jours. Elle comporte actuellement plus de 12 000 *exempla* indexés couvrant un long Moyen Âge s'étendant de l'Europe latine à Byzance et l'Asie, provenant de sources chrétiennes, juives, islamiques et bouddhistes. Chaque *exemplum* est soigneusement encodé, accompagné de métadonnées et relié à des mots-clés et des thésaurus, afin de permettre aux chercheurs de lancer une grande variété de requêtes au sein de ce corpus d'artéfacts culturels. La plus récente transformation de ThEMA a été sa conversion au format XML-TEI au sein d'une base de données offrant de puissantes interfaces qui ont permis au

projet de s'engager entièrement dans les standards des « FAIR data », afin d'assurer longévité et usage élargi des données pour les chercheurs et le grand public.

Medieval sermons brim with tens of thousands of *exempla*, the “exemplary short stories” frequently used to convey belief and morals to the faithful. Long considered a class of dogmatic, religious, and moral texts, they were mostly studied as a subgroup of folktales and fables in literary studies. With the “anthropological turn” in historical research, new attention was paid to *exempla* as sources of cultural and historical insight, notably by the historian Jacques Le Goff and his student-inheritors.

ThEMA, a database of medieval *exempla*, began life in the 1990s in the hands of these inheritors, and has grown and transformed since then. It currently holds over 12,000 indexed *exempla* from the long “global Middle Ages”, stretching from Latin Europe to Byzantium to Asia, and drawing from Christian, Jewish, Islamic, and Buddhist sources. All *exempla* are carefully encoded and assigned metadata, and linked to keywords and thesauri, to enable researchers to undertake a variety of inquiries into these cultural artifacts. The most recent transformation of ThEMA into XML-TEI format, within a searchable database that provides powerful interfaces, has permitted the project to fully engage with FAIR data standards to ensure longevity and broader usability of the data for researchers and the general public.

Entrées d'index

MOTS-CLÉS : histoire médiévale, littérature, sciences des religions, TEI, base de données

KEYWORDS: medieval history, literature, religious studies, TEI, database

Introduction

¹ La prédication médiévale et la littérature didactique comportent des dizaines de milliers d'*exempla* ou « brèves histoires exemplaires », souvent utilisées pour transmettre doctrine et morale religieuses aux fidèles. Malgré le travail pionnier des historiens depuis le début du ^{xx}^e siècle pour cataloguer la masse de ces *exempla* (Welter 1973), ces sources ont longtemps été considérées comme peu intéressantes du point de vue littéraire et d'un bas niveau concernant leur portée dogmatique et morale sur le plan religieux.

² C'est l'historien Jacques Le Goff et son groupe d'étudiants et disciples, recourant à l'anthropologie historique de l'École des Annales, qui ont attiré de nouveau l'attention sur ces histoires comme sources d'information sur les conceptions du monde des médiévaux, en posant la question suivante : pourquoi et comment les clercs médiévaux ont-ils fait

confiance au *storytelling*, particulièrement au récit exemplaire (*exemplum*), pour prêcher aux fidèles (Bremond, Le Goff et Schmitt 1996 ; Berlioz 2007, 297-312) ?

3 Un certain nombre d'héritiers de Jacques Le Goff ont depuis donné suite à ses intuitions initiales au travers de recherches plus structurées dans la masse des *exempla* et des données qu'ils contiennent. Dans les années 1980, un groupe de travail dirigé par Jacques Berlioz et Marie-Anne Polo de Beaulieu, dans le cadre d'un séminaire à l'EHESS, fit les premiers essais d'indexation des *exempla*, en pensant d'abord à simplement prolonger l'*Index exemplorum* de Frederic C. Tubach (1969) – à cette époque le catalogue le plus abouti de récits tirés des *exempla* médiévaux. Ce travail initial évolua rapidement vers une relecture critique aboutissant à des corrections de ce dernier (Berlioz et Polo de Beaulieu 1992).

4 En 1932, Dean Stith Thompson avait proposé des motifs narratifs classés par thèmes. Il leur avait rattaché certains *exempla* médiévaux (*Sermones vulgares* de Jacques de Vitry, *The Exempla of the Rabbis*, *Gesta Romanorum*, etc.), mais ils se trouvaient noyés dans des références à des récits européens et extra-européens. Pour reprendre J. Berlioz, « Les motifs strictement chrétiens sont peu nombreux et intégrés dans une thématique large (*Magic, Marvels*, etc.). Et placer des *exempla* sous des thèmes trop généraux s'avère peu utile » (Berlioz 2005). Les mêmes problèmes se rencontrent dans l'index des contes-types de Antti Aarne et Stith Thompson (1961). Dans un domaine beaucoup plus spécialisé, F. C. Tubach reprend la même méthodologie : regrouper 44 000 variantes d'*exempla* sous 5 400 types de récits exemplaires dans son *Index exemplorum* paru en 1969. Les rattachements sont en général pertinents mais il existe de notables exceptions, et cette méthodologie efface les variantes propres à chaque *exemplum*, alors qu'elles sont porteuses de beaucoup d'informations, pour l'historien comme pour le spécialiste de la littérature.

5 C'est pourquoi le groupe ThEMA, après avoir pensé faire des ajouts et des corrections à l'*Index exemplorum* de Tubach, s'est tourné vers une méthodologie radicalement opposée : prendre chaque *exemplum* comme un objet culturel unique et l'indexer individuellement. Cependant, ThEMA établit des liens en utilisant une grille multidimensionnelle pour interroger et indexer les *exempla* (Polo de Beaulieu et Collomb 2005, Berlioz 2016). L'accent mis sur un *exemplum* donné consiste à analyser ses dimensions historiques et anthropologiques. Cette enquête thématique révisée s'est formée autour des catégories fondamentales proposées par Gérard Blangez dans son *Repertorium exemplorum* (gracieusement offert à ThEMA). Elle s'est ajoutée à la grille très fine proposée par Alain Boureau (1984) lors du symposium d'Oxford pour l'étude des sermons médiévaux, qui avait orienté sa grille sur le rôle des auteurs de recueils d'*exempla* et le contexte dans lequel ils ont écrit.

6 Ces développements théoriques se sont matérialisés, dans les années 1990, par les séminaires susmentionnés, en un ensemble normalisé de formulaires enregistrés dans des modèles Microsoft Word. Il s'agissait des premiers enregistrements de données pour le projet ThEMA. Au niveau du « memento » concernant le recueil, ils comprenaient des informations sur l'auteur d'une collection donnée d'*exempla*, le texte source sur lequel s'appuie l'indexation (manuscrit, édition imprimée, etc.) et des références bibliographiques. Au niveau de chaque *exemplum* de la

collection, une grille d'indexation donnait une référence standardisée, le situait dans le recueil (« contexte »), fournissait un résumé, des mots-clés, une ou des sources et des textes parallèles, un renvoi à l'*Index exemplorum* de Tubach et à d'autres catalogues, une bibliographie et un commentaire. En bref, la méthodologie de ThEMA commence par un *exemplum* lui-même, situé dans la collection de l'auteur, en créant à partir de là des relations avec d'autres *exempla*, avec d'autres sources et dans l'historiographie.

7 Le projet ThEMA est né d'une collaboration issue des séminaires de l'EHESS qui ont permis aux chercheurs d'échanger leurs idées sur l'indexation et de développer un modèle, avec l'aide importante de Marjorie Burghart (CNRS-CIHAM), qui avait développé le site original en PHP et MySQL (Burghart 2005). Cette collaboration a permis à de nombreux chercheurs d'apporter, au cours des vingt dernières années, de riches contributions au projet – à ce jour, plus de 50 contributeurs en cinq langues ont indexé plus de 12 000 *exempla*. Les frontières géographiques ont été étendues et comprennent maintenant, entre autres, la Russie médiévale, la Chine et le Japon, ce qui élargit le champ des comparaisons¹. La base matérielle de ThEMA s'est également élargie pour inclure des images étroitement liées aux *exempla*, ainsi que l'interconnexion avec d'autres bases de données, par exemple en reliant les exemples de ThEMA aux sermons de Sermones.net (notamment pour les *exempla* issus des sermons de Jacques de Voragine).

8 Avec la croissance du projet et le partage des données de ThEMA entre les collaborateurs et les chercheurs intéressés, une série de changements est intervenue dans la collecte et la gestion des données. Vers 2002, les données des documents Microsoft Word originaux ont été transférées dans une base de données MySQL par M. Burghart avec une interface de saisie sous Microsoft Access qui permettait le travail des indexateurs hors ligne, dans un environnement familier. Cette interface a été à son tour remplacée en 2005 par un site Web directement lié à une base de données MySQL pour permettre une collaboration dynamique dans la collecte et le contrôle de la qualité des données, ainsi que pour permettre l'accès à un public plus large. Ce dernier site a été lancé grâce au travail de M. Burghart et d'un troisième chef de projet intégré à l'équipe, Pascal Collomb (EHESS). En 2015, M. Burghart a proposé de transférer le projet ThEMA à Huma-Num, la très grande infrastructure de recherche (TGIR) pour les sciences humaines financée par le CNRS, afin de sécuriser l'environnement du serveur à long terme. Il a ensuite été décidé d'élaborer un schéma XML-TEI pour les données de ThEMA et de les verser dans une base de données et serveur d'application eXist-db, sous la direction de Jean-Paul Rehr (université Lumière Lyon 2 et CIHAM), outil lancé à l'automne 2019².

Méthodes

Sélection des sources, méthodologie de l'indexation

⁹ Le projet s'intéresse à des *exempla* du Moyen-Âge, une période définie approximativement comme allant de 500 à 1500 après J.-C. La plus grande partie de la base de données provient de sources latines rédigées entre 1100 et 1350, période de la plus intense production d'*exempla*, en particulier par les ordres cistercien et dominicain. Cette tendance est liée à l'accent mis par les monastères sur la collecte de connaissances et le « folklore monastique » (McGuire 2010, 422), dans le premier cas, et à l'accent mis sur la prédication au peuple, dans le second. Cependant, ThEMA ne se concentre pas exclusivement sur l'Europe latine médiévale – pour ne prendre que quelques exemples, les travaux des compilateurs bouddhistes japonais médiévaux Kamo no Chomei et Minamoto no Takakuni, ainsi que le *Sefer Hamaassiyot* anonyme en hébreu, permettent aux chercheurs de déployer une analyse comparative. Nonobstant les aires culturelles d'origine de ces compilations, elles sont soumises à la même méthodologie d'indexation.

¹⁰ Le modèle d'organisation de ThEMA repose sur les collections (ou recueils) : des compilations d'*exempla*, souvent des centaines, écrits ou compilés par un auteur au Moyen Âge. Le travail d'indexation commence par la sélection d'une collection qui intéresse le chercheur et l'équipe de ThEMA et qui a déjà été publiée dans une bonne édition de référence (bien qu'à l'occasion les membres de l'équipe de ThEMA indexent des documents manuscrits ou des éditions imprimées anciennes). Pour les besoins de ThEMA, une collection est « intéressante » si elle rassemble une variété d'histoires nouvelles et anciennes, insérées dans un commentaire ou un thème élaboré par le compilateur médiéval lui-même et qui détermine le processus de sélection des récits exemplaires par l'auteur. Les *exempla* s'avèrent être de puissants révélateurs, qui opèrent sur de multiples niveaux de signification historique et de transmission orale ou écrite, latine ou vernaculaire.

¹¹ Selon l'édition utilisée pour l'indexation, l'indexateur ou l'indexatrice fournit un résumé de chaque *exemplum* ou cite le résumé de l'édition si celui-ci est disponible (si l'édition est dans le domaine public, ou si l'éditeur le permet). L'indexateur procède ensuite à l'identification de la source ou des sources de l'histoire qui est au cœur de l'*exemplum*. Un bon compilateur médiéval aura fourni les sources de ses *exempla*, ce qui peut faciliter le travail de l'éditeur et de l'indexateur. Cependant, ce dernier doit encore parfois faire la distinction entre les sources revendiquées par le compilateur et la source réellement utilisée par lui : souvent, les compilateurs médiévaux se réclament de l'autorité et du prestige en revendiquant des sources anciennes ; en vérité, beaucoup d'*exempla* sont issues de compilations intermédiaires, qui laissent leurs traces. Il s'agit souvent d'un travail minutieux pour l'éditeur et l'indexateur, et non d'un simple exercice intellectuel : l'examen et la vérification des sources d'un *exemplum* permettent aux chercheurs de comprendre les méthodes de travail et le milieu de l'auteur du recueil, ainsi que les pratiques de trans-

mission de la culture orale ou écrite à une époque pour laquelle nous disposons de peu de preuves directes (Polo de Beaulieu et Berlioz 2005 ; Collomb 2007 ; Polo de Beaulieu et Collomb 2009).

12 Les compilations d'*exempla* médiévales sont fréquemment présentées par leur auteur sous la forme d'un catalogue, éventuellement classé selon l'ordre alphabétique, par exemple l'*Alphabetum narrationum* (« Alphabet des récits ») d'Arnold de Liège, ou la *Scala coeli* (« Échelle du ciel ») de Jean Gobi, ce qui permet de les transposer relativement facilement dans une base de données. Parfois, cependant, les récits exemplaires sont intégrés dans des dialogues (sur le modèle ancien des *Dialogues* de Grégoire le Grand), comme ceux de la *Disciplina clericalis* (« Discipline de clergie ») de Pierre Alphonse ou du *Dialogus miraculorum* (« Dialogue des miracles ») de Césaire de Heisterbach. On les trouve aussi dans des ouvrages comme le volumineux *Tractatus de diversis materiis praedicabilibus* (« Traité des diverses matières à prêcher ») de l'inquisiteur dominicain Étienne de Bourbon et les sermons de Bernardin de Sienna. Dans ces cas, l'indexateur doit extraire les *exempla* de manière raisonnée et décider s'il convient d'inclure le commentaire et les citations qui les entourent. Certaines compilations peuvent également être traduites en langues vernaculaires, à partir d'un original en latin et, à ce titre, l'édition vernaculaire sera indexée à part, car les traductions médiévales sont fréquemment l'occasion de révisions et de remaniements, comme c'est le cas pour les traductions anglaise et catalane de l'*Alphabet des récits*.

13 Parce que la méthodologie de l'indexation est fondée sur l'unicité de chaque *exemplum*, les *exempla* ne sont pas regroupés par type d'histoire, à la manière de l'*Index exemplorum* de F. C. Tubach. Chaque *exemplum* est plutôt relié aux autres par une série de caractéristiques communes : tout d'abord, les mots-clés fondés sur un vocabulaire ouvert. Les mots-clés ne résument pas un *exemplum* donné et ne sont pas non plus tirés de celui-ci, mais sont plutôt dérivés des concepts illustrés par l'*exemplum* (l'index actuel des mots-clés compte environ 2 200 items). Par exemple, si un *exemplum* décrit la confession de péchés à un prêtre, les mots-clés « confession », « sacrement », « rituel » permettent de relier cet *exemplum* à d'autres qui partagent les mêmes concepts. Avec le texte source et le résumé de l'*exemplum*, les mots-clés permettent de réfléchir à l'influence d'un auteur sur un autre. La deuxième caractéristique commune est constituée par les références à d'autres index, facilitées par les systèmes de numérotation de l'*Index exemplorum* de Tubach et des typologies d'Arne et Thompson. Ceux-ci permettent de passer d'une typologie et d'un concept à l'autre. Une autre couche d'information est apportée par la bibliographie propre à chaque *exemplum* (éditions, études, articles d'encyclopédies et de dictionnaires enregistrés dans un format cohérent), permettant à l'utilisateur de parcourir les *exempla* via leur bibliographie.

14 Voyons ci-dessous un exemple d'indexation d'un récit exemplaire, le n^o 538 de la *Scala coeli* (« Échelle du ciel ») de Jean Gobi :

- Référence : *Scala coeli*, n^o 538 [éd. Polo de Beaulieu, p. 399-400 et 685].
- Contexte : Rubrique « Des fils ».
- Résumé : Un roi malade envoie ses trois fils en quête de l'eau de jeunesse seule capable de le guérir. Les deux plus âgés passent par les régions les plus accessibles tandis que le benjamin passe par les

montagnes. C'est lui qui trouve – à l'issue d'une série d'épreuves – l'eau de jouvence. Il se marie avec une belle princesse, revient et hérite du royaume. Lecture allégorique du récit : la montagne symbolise la pénitence, la princesse, la charité, l'eau de jouvence, l'eau de la rémission des péchés, etc.

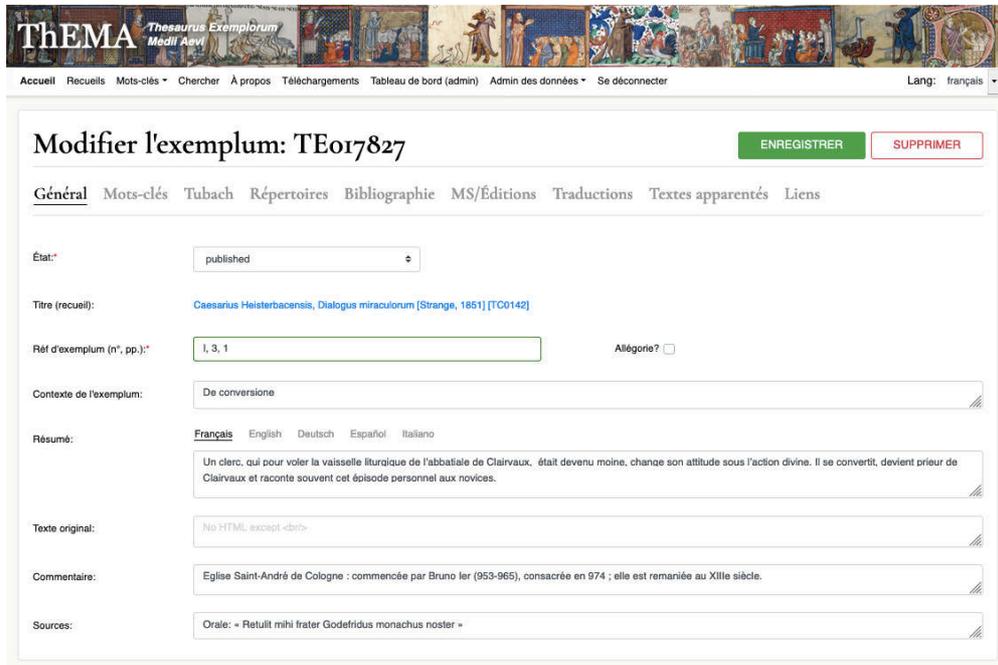
- Allégorie : OUI.
- Sources : pas de source repérée.
- Répertoires : Conte-type Aarne-Thompson 551 ; Tubach 3638.
- Bibliographie : Joël H. Grisward, « Le voyage du fils. Structuration de l'espace et itinéraire trifonctionnel dans un *exemplum* de la *Scala coeli* », dans *Romania*, t. 117, 1999, p. 545-555.
- Mots-clés : parenté, maladie, épreuve, mariage, héritage, jeunesse.

Collecte des données et contrôle de leur qualité

15 Fondé comme un projet collaboratif, ThEMA s'est concentré depuis le début sur des outils destinés à créer des liens entre les données enregistrées et qui assurent un certain niveau de qualité scientifique à ces données. À l'origine, cette tâche était assurée par des modèles Microsoft Word, puis par une base de données MySQL en ligne (qui réutilisait les descriptions de champs des modèles) avec un accès restreint à la saisie des données. En 2019, le corpus de ThEMA a été soumis à une transformation complète (près de 12 000 *exempla* plus les 2 200 mots-clés et la bibliographie) au format XML-TEI, dans une base de données remaniée dans eXist-db. La plupart des méthodes de gestion du contrôle de la qualité dans les systèmes précédents ont été reprises dans le nouvel environnement, mais elles ont été automatisées pour les tâches les plus simples, et davantage d'outils ont été fournis pour permettre à de nouveaux chercheurs d'élargir le contenu de la base de données.

16 Les chercheurs ne créent pas directement les fichiers XML-TEI. Ils utilisent plutôt une série de masques de saisie ou d'écrans sur le site Web de ThEMA pour interagir avec la base de données (voir figure 1). L'accès à ces masques de saisie est basé sur des autorisations authentifiées. Un membre de l'équipe ThEMA se voit attribuer un « rôle » concernant soit l'ensemble de la base de données, soit un recueil spécifique. Une poignée de membres de l'équipe qui ont un accès administrateur à la base de données peuvent gérer les utilisateurs et tous les documents du système. La plupart des membres de l'équipe ont un rôle d'indexateur ou d'éditeur. Un éditeur décide quels recueils sont ajoutés à la base de données, et gère et contrôle le travail des indexateurs avant la publication. Cette séparation des rôles est utile en particulier pour les chercheurs qui souhaitent faire appel à des jeunes collaborateurs (tels que des étudiants), dont les travaux peuvent ainsi être révisés, tout en limitant l'impact des erreurs dans la base de données. Quoi qu'il en soit, tant les indexateurs que les éditeurs ne sont autorisés à modifier que les recueils spécifiques qui leur ont été attribués.

FIGURE 1. CAPTURE D'ÉCRAN POUR L'INDEXATION D'UN *EXEMPLUM*

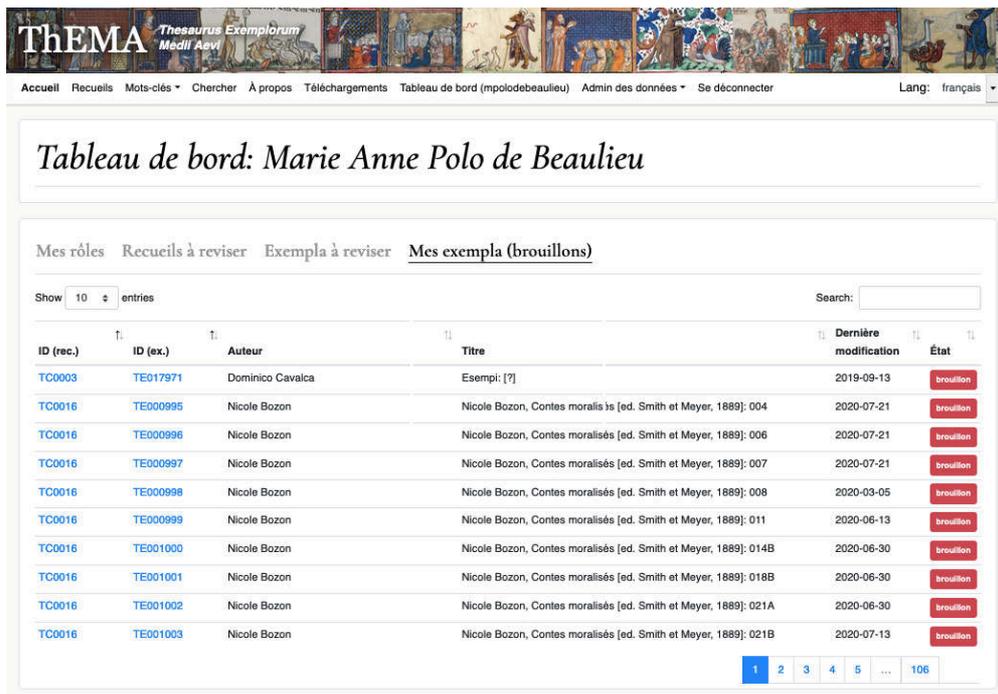


17

La base de données ThEMA gère le processus d'indexation à travers quatre statuts de documents que le responsable adapte à chaque utilisateur dans un tableau de bord (voir figure 2) :

1. brouillon : un *exemplum* est créé et se trouve en cours de modification
2. soumis : un *exemplum* a été soumis à un éditeur pour approbation
3. approuvé : un *exemplum* a été approuvé par un éditeur
4. publié : un *exemplum* approuvé a été publié, il est désormais accessible au public

FIGURE 2. CAPTURE D'ÉCRAN DU TABLEAU DE BORD POUR GÉRER LES TÂCHES D'INDEXATION



18 Les masques de saisie des données eux-mêmes contiennent une variété d'outils pour assurer la cohérence et la connexion des données à travers la base de données. Pour décrire un recueil dans son ensemble, le memento contient le nom de l'auteur, le lieu de production, les affiliations (dominicaine, bénédictine, etc.), qui ne peuvent être sélectionnés qu'à partir d'index propres à ThEMA et contrôlés par le biais de listes déroulantes ; les dates sont validées pour s'assurer qu'elles sont conformes à un format ISO commun.

19 Pour chaque *exemplum*, les éléments suivants ne peuvent être ajoutés qu'au moyen de listes déroulantes :

- mots-clés conceptuels : une liste approuvée de mots-clés gérée par l'administrateur
- références Tubach : une liste des numéros Tubach gérée dans la base de données
- références bibliographiques : les utilisateurs ne peuvent faire leur choix qu'à partir d'une bibliographie qui a été ajoutée au groupe Zotero de ThEMA (Bibliex, voir ci-dessous)

20 Lors de la saisie, les données sont validées d'abord par des scripts dans le navigateur (pour les champs obligatoires manquants) et ensuite sur le serveur lorsque l'utilisateur tente de les sauvegarder. Une erreur de validation produit un message d'erreur détaillé à l'intention des utilisateurs pour les aider à corriger les problèmes. Toutes les modifications apportées aux données sont enregistrées avec le nom de l'utilisateur, la description de l'action et un horodatage.

Encodage des données

TEI

21 Vers le milieu de l'année 2019, toutes les données textuelles de ThEMA ont été converties au format XML selon le schéma TEI P5 et toutes les données depuis lors sont encodées de cette manière. Le XML-TEI est un schéma d'encodage de textes savants développé et maintenu par le consortium Text Encoding Initiative³ et reste le format le plus courant au niveau international pour l'encodage des données textuelles et des métadonnées dans les sciences humaines. La version spécifique ThEMA du schéma TEI P5 est disponible en ligne⁴.

22 Le projet utilise la structure de corpus de la TEI, avec un recueil stocké dans un élément <teiCorpus> et chaque *exemplum* stocké dans un élément <TEI>. Un document relatif à un recueil intègre les documents *exempla* via l'élément <xi:include>. Comme une grande partie du contenu produit par le groupe ThEMA consiste en des métadonnées (c'est-à-dire des descriptions des *exempla*), la majeure partie des données se trouve dans l'élément <teiHeader> de chaque document, en particulier l'élément <sourceDesc> pour la description de l'*exemplum* et sa bibliographie, et l'élément <profileDesc> pour les mots-clés. En effet, les seules données qui se trouvent dans l'élément <text> sont les résumés des *exempla* dans l'une des cinq langues possibles du groupe de travail (français, anglais, allemand, espagnol, italien). Toutes les références partagées entre les *exempla* – auteurs, lieux, affiliations ou contextes,

mots-clés et groupes de mots-clés – sont stockées dans des fichiers TEI séparés (accessibles les uns aux autres grâce à l'utilisation des éléments <xi:include> et <standOff>).

Mots-clés

23

Il convient de mentionner les index des mots-clés de ThEMA (fichier « list_keywords.xml ») et des groupes de mots-clés (fichier « list_keywordgroups.xml »), qui offrent un accès riche au contenu thématique. Les quelque 2 200 mots-clés, contrôlés par l'équipe éditoriale de ThEMA, constituent le premier point d'indexation et souvent le premier lieu d'exploration pour les chercheurs. Les « groupes de mots-clés » ont été ajoutés en 2020 comme méthode d'exploration de plus haut niveau des mots-clés et des *exempla* (voir le tableau 1). Les relations entre les groupes de mots-clés et les mots-clés sont multivalentes – un seul mot-clé peut appartenir à plusieurs groupes, de sorte que le mot-clé « meurtre » peut appartenir aux groupes « pouvoir » et « morale/éthique ». Les mots-clés et les groupes de mots-clés font partie de l'aspect multilingue du projet et sont représentés dans les cinq langues du groupe de travail.

TABLEAU 1. GROUPES THÉMATIQUES DE MOTS-CLÉS (DANS DEUX DES CINQ LANGUES DU SITE) AVEC LE NOMBRE DE MOTS-CLÉS ASSOCIÉ À CHAQUE GROUPE

ID	Français	English	Mots-clés
TH01	église	<i>church</i>	245
TH02	pouvoir	<i>power</i>	153
TH03	vie quotidienne	<i>daily life</i>	287
TH04	arts et savoirs	<i>arts/sciences</i>	91
TH05	au-delà	<i>afterlife</i>	53
TH06	élément naturel	<i>natural elements</i>	105
TH07	vie morale	<i>morals/ethics</i>	292
TH08	Bible	<i>Bible</i>	87
TH09	corps	<i>body</i>	169
TH10	vie religieuse non chrétienne	<i>non-Christian religious life</i>	21
TH11	animaux	<i>animals/creatures</i>	92
TH12	noms de personne	<i>person name</i>	372
TH13	noms de lieux	<i>place name</i>	124

Connexion avec les ontologies publiques

24

ThEMA est progressivement connecté avec des ontologies publiques. Les premières tentatives ont été faites avec l'index des auteurs en utilisant les ontologies de référence du VIAF⁵ et de la Bibliothèque nationale de France ARK. Il en sera de même pour les lieux et les affiliations utilisant les vocabulaires Getty. La grande diversité des listes de mots-clés – qui témoigne de l'étendue et de la richesse des *exempla* – a empêché de se connecter aux ontologies publiques de manière cohérente et utile. C'est pourquoi le groupe ThEMA envisage de créer une ontologie publique qui serait réutilisée par d'autres projets, en 2022.

Zotero

25 La base de données bibliographiques de ThEMA est gérée dans un groupe Zotero⁶, qui met les données à la disposition du public dans des formats réutilisables séparément. Ces données bibliographiques sont continuellement synchronisées avec la base de données de ThEMA via une API et stockées sous forme de fichiers TEI dans un élément <listBibl>. L'identifiant Zotero fonctionne comme le @xml:id dans la base de données ThEMA et relie toutes les références bibliographiques. Cette connexion à Zotero et le contrôle de la qualité par ce moyen sont une nouveauté de la version XML de ThEMA. Alors que toutes les nouvelles références bibliographiques doivent désormais passer par Zotero, les plus de 20 000 notices bibliographiques de l'ancienne base de données ThEMA ont été systématiquement converties en références liées à Zotero (à la date de publication, 85 % des références ont été liées à Zotero grâce à un programme de nettoyage des données).

Références Tubach

26 Étant donné l'importance historique de l'*Index exemplorum* de Frederic Tubach pour l'étude des *exempla*, le groupe de travail ThEMA a entrepris la création d'une liste de références des numéros de Tubach au format XML. Cette liste, utilisable uniquement dans la base de données ThEMA, garantit que toute référence à Tubach est cohérente et bien structurée pour les requêtes. Les utilisateurs peuvent ainsi faire des allers-retours entre les références ThEMA et Tubach.

Métadonnées Dublin Core

27 Dans tous ses fichiers TEI, ThEMA utilise les métadonnées du Dublin Core pour la description des documents, stockées dans l'élément de TEI <xDATA>. Ces données sont générées automatiquement par la base de données et sont utilisées pour l'affichage de la notice sur le site Web de ThEMA (dans l'élément HTML <head>) et pour le catalogage des données avec Nakala, partenaire de ThEMA pour les archives (voir ci-dessous). ThEMA utilise les éléments suivants du Dublin Core :

- *identifier* : soit « <http://thema.huma-num.fr/collection/TC----> » (collection), soit « <http://thema.huma-num.fr/exempla/TE-----> » (*exempla*), où « ---- » représente l'identifiant correspondant de la base de données ThEMA
- *title* : titre de la collection ou de l'*exemplum*, issu de l'élément TEI <titleStmnt>
- *publisher* : « <http://thema.huma-num.fr> »
- *editor* : les noms des éditeurs et indexateurs qui travaillent sur le document
- *rights* : « <http://creativecommons.org/licenses/by-nc-sa/4.0/> »
- *type* : « <http://purl.org/dc/dcmitype/Dataset/> »
- *relation* : « <http://thema.huma-num.fr/collection/TC----> » où « ---- » représente l'identifiant du recueil auquel l'*exemplum* est lié (élément utilisé uniquement pour les *exempla*)

Données non textuelles

28 Dans le but d'examiner la signification d'un *exemplum* dans son contexte, certains contributeurs incluent des images liées à l'*exemplum*. Il s'agit souvent d'images de manuscrits enluminés, mais également des peintures murales ou des gravures, ce qui ajoute une dimension culturelle supplémentaire au texte. Ces images sont décrites en relation avec le texte et liées au fichier XML-TEI correspondant.

Archivage des données

29 La transformation de la base de données ThEMA en XML-TEI (par opposition à un schéma personnalisé) faisait partie d'un plan visant à assurer la longévité et la réutilisation de ses données. Ce changement important n'aurait pas pu être réalisé aussi facilement sans le soutien de la TGIR Huma-Num, la « très grande infrastructure de recherche » en sciences humaines et sociales financée par le CNRS⁷. Huma-Num fournit à ThEMA deux moyens de s'assurer que le projet peut mettre en œuvre des pratiques de données FAIR⁸.

30 Le premier est la base de données et le site Web de ThEMA. Huma-Num fournit l'infrastructure technique sur laquelle repose la base de données de ThEMA (dans eXist-db), stocke et met les données à la disposition du grand public. Sauf fermeture improbable d'Huma-Num et de ses services, ou refus de soutenir certains projets, il n'y a aucune raison de penser que les données de ThEMA ne seront pas accessibles au public à long terme sur <http://thema.huma-num.fr>. À cela s'ajoute la décision de conception prise pour ThEMA : le déploiement d'eXist-db résulte d'un choix réfléchi d'utiliser une infrastructure qui n'a pas de dépendances extérieures. Aucune modification du serveur ne peut affecter le fonctionnement de l'environnement d'application autonome fourni par eXist-db ; cela contraste avec les environnements qui utilisent PHP, Python ou d'autres ensembles de code qui dépendent habituellement de l'administrateur du serveur et des programmeurs, qui doivent être continuellement en phase avec les mises à jour de versions. En pratique, les seuls risques pour le fonctionnement à long terme de la base de données et du site de ThEMA sont soit le retrait soudain des services de Huma-Num, soit l'incapacité du groupe de ThEMA à surveiller l'instance eXist-db pour résoudre des bugs inattendus. En tout état de cause, eXist-db fournit des sauvegardes automatisées des données dans des formats permettant d'effectuer rapidement la restauration dans une autre instance de la base de données.

31 Pour contrebalancer ces risques, ThEMA a une seconde stratégie pour l'archivage des données à long terme, qui repose sur des identifiants pérennes : une division distincte d'Huma-Num, Nakala, fournit « un entrepôt sécurisé qui assure à la fois l'accessibilité aux données et leur citabilité dans le temps ». Ce service de stockage numérique sécurisé à long terme, garanti par sa relation avec le CNRS, fournit des identifiants permanents ainsi que l'archivage des métadonnées. Les identifiants permanents sont des DOI, ce qui permet à ThEMA de disposer d'un mécanisme de citation robuste⁹. Les métadonnées sont stockées selon les normes OAI-PMH (Open Archives Initiative Protocol for Metadata Har-

vesting) et sont basées sur le Dublin Core (en utilisant les métadonnées Dublin Core de ThEMA, voir ci-dessus) ; elles sont mises à la disposition du public par Huma-Num de diverses manières, y compris par des triplets RDF. Les données ThEMA sont archivées avec Nakala sous forme de fichiers TEI compressés au format ZIP, chaque fichier contenant une collection – les DOI dans ThEMA représentent donc des recueils. En cas de modification des données dans la base de données ThEMA, Nakala est automatiquement mis à jour avec les données modifiées.

Réutilisation de la base de données ThEMA

Accessibilité des données

- 32 Les données de ThEMA sont mises à disposition de quatre façons :
1. Lire et chercher les données directement sur le site
 2. Télécharger les fichiers XML-TEI depuis le site
 3. Télécharger les recueils et les *exempla* au format PDF depuis le site
 4. À partir du printemps 2022, télécharger les fichiers XML-TEI depuis la plateforme Nakala

Potentiel de recherche

- 33 Les *exempla* fournissent un riche matériel aux historiens de la vie intellectuelle, culturelle et sociale, ainsi qu'aux spécialistes de la littérature et de la religion. Ces récits brefs et moraux saisissent des dynamiques culturelles complexes, car les *exempla* contiennent un mélange de sources orales (expérience directe de l'auteur, contes, proverbes, refrains, secrets de confession et aveux devant l'inquisition) et de sources écrites, allant de la doctrine (théologie, exégèse) aux instruments de travail (recueils alphabétiques des thèmes bibliques appelés *distinctiones* bibliques, concordances bibliques, recueils d'*exempla* des confrères, encyclopédies, bestiaires), en passant par les chroniques et les œuvres de fiction. En reliant des *exempla* à des sources plus anciennes, nous pouvons voir comment, par exemple, l'héritage des fables animalières antiques (les fables d'Ésope) a été transmis à travers le temps et retravaillé au Moyen Âge dans les sermons.

- 34 En même temps, les *exempla* ont été considérés comme une sorte de « bible de la vie quotidienne » (Berlioz et Polo de Beaulieu 2002) pour les nombreux aperçus de la vie quotidienne qu'ils donnent. Jacques Berlioz a montré comment, dans le vaste ouvrage d'Étienne de Bourbon, la peur était inculquée aux enfants par la figure du croquemitaine (« faire Barbo ») (Alexandre-Bidon et Berlioz 1998). La lumière est également faite sur les origines d'histoires autrefois difficiles à dater comme le Petit Chaperon rouge (Berlioz 1991), ou de mythes comme celui de Don Juan (Polo de Beaulieu 1991, 488, 705). Le contexte de production de nombreux *exempla* – les monastères et les couvents médiévaux – se retrouve naturellement dans les récits, révélant des scènes intimes de la vie dans les

cloîtres rarement documentées autrement : les relations entre novices, convers et moines de chœur, les tensions entre l'abbé et ses moines, les rituels et la vie quotidienne.

35 En opposition à l'image dominante du Moyen Âge latin comme société chrétienne immuable, les *exempla* révèlent des interactions mouvantes entre l'Église et la société. À travers ces récits, nous voyons la transmission du catéchisme, la lutte contre la superstition et l'hérésie telles qu'elles étaient perçues, la promotion de la confession – autant de nouveautés liturgiques récentes qui s'inscrivent dans des conflits plus importants sur le contrôle et la centralisation des croyances (quatrième concile du Latran, 1215). Malgré l'évidence de ces grands changements dans la société, on peut identifier, grâce aux *exempla*, des transformations locales plus fines dans la structuration des croyances et des pratiques, comme l'invention liturgique du lundi des Trépassés, diffusée via des récits exemplaires, dès le XI^e siècle (Polo de Beaulieu 1998).

36 Avec la récente transformation de la base de données ThEMA en un balisage TEI très lisible, les recherches qui reposaient auparavant sur des enquêtes qualitatives sur les *exempla* peuvent maintenant être enrichies et remises en question par des analyses quantitatives. Combinées aux outils mis à disposition par le site Web, les hypothèses peuvent être testées plus rapidement et de manière plus approfondie.

Limites et lacunes

37 Une limitation importante des données de ThEMA est l'absence des textes sources dans la base de données. Cela est généralement dû aux limitations des droits d'auteur sur les éditions scientifiques utilisées pour l'indexation des *exempla*. Le projet a tenté de combler cette lacune au moyen de résumés fournis par les indexateurs. Ces résumés sont eux-mêmes multilingues, ce qui reflète la diversité des contributeurs de ThEMA ; bien que les médiévistes travaillent généralement dans plusieurs langues, les chercheurs qui ne sont pas familiers d'un contexte de recherche multilingue pourront rencontrer quelques difficultés. Ici, la bibliographie solide et bien documentée de ThEMA aide l'utilisateur à retrouver les textes sources et même, le cas échéant, des éditions dans différentes langues.

38 Bien que la base de données ThEMA impose un cadre pour l'indexation des *exempla* et un processus de révision qui vise à donner un minimum d'assurance sur la qualité des données, l'indexation n'est pas un exercice strictement objectif. L'évaluation qualitative de la signification d'un *exemplum* et du contexte qui l'a produit est finalement laissée aux chercheurs spécialisés auxquels le projet a fait confiance. Toute recherche d'*exempla* à grande échelle à partir des données de ThEMA doit tenir compte d'une certaine variabilité dans les approches des contributeurs.

Contributions futures

39 ThEMA n'est pas un projet à durée déterminée. La collaboration s'enrichit continuellement des contributions de nouveaux chercheurs qui ajoutent de nouveaux recueils et des étudiants qui apprennent à indexer sous la direction de chercheurs expérimentés. Pendant ce temps, l'équipe centrale continue d'ajouter du matériel et d'améliorer les contributions passées.

40 En 2021, ThEMA a vu l'intégration de certains textes sources issus du domaine public (*Dialogus miraculorum* et *Libri VIII miraculorum* de Césaire de Heisterbach et *Scala coeli* de Jean Gobi) : ces éditions complètes sont mises à disposition à côté des *exempla* indexés afin d'offrir une expérience plus riche à l'utilisateur¹⁰. En 2022, le groupe expérimentera également l'établissement de liens avec des ontologies publiques dans le cadre de ses pratiques de données FAIR.

Éthique et droits

41 Toutes les données de ThEMA sont historiques, toutes les images sont du domaine public et, en tant que telles, aucun problème éthique concernant la confidentialité des données n'est à signaler.

42 Toutes les contributions de données textuelles à ThEMA sont mises à la disposition du public sous la licence Creative Commons « Attribution – Pas d'utilisation commerciale – Partage dans les mêmes conditions » (Attribution-NonCommercial-ShareAlike) 4.0. Les images ont des droits variables, déclarés pour chaque image. Lorsque les images sont produites par l'indexateur, elles sont mises à disposition sous la licence Creative Commons susmentionnée. Si une image a été obtenue auprès d'un dépôt d'archives, d'un musée ou d'une autre institution, la déclaration des droits de ThEMA suit les autorisations et les politiques de l'institution.

Sources de financement

43 Durant les vingt dernières années, le projet ThEMA a reçu des fonds de l'ANR (projet « Une culture du Livre dans une société d'illettrés. Parler, figurer, écrire dans l'Europe médiévale », 2006-2010), de l'EHESS, du CNRS, du CIHAM, du Labex HASTEC, du CRH (EHESS-CNRS), de SAPRAT (EPHE), de l'IRHT (CNRS), du Césor (EHESS) et du Consortium COSME (Huma-Num).

Bibliographie

Aarne, Antti et Stith Thompson. 1961. *The Types of the Folktale : a Classification and Bibliography*. 2^e éd. Helsinki : Academia Scientiarum Fennica.

Alexandre-Bidon, Danièle et Jacques Berlioz. 1998. « Le masque et la barbe. Figures du croquemitaine médiéval ». *Le Monde alpin et rhodanien*, 26 (2-4 : « Les croquemitaines. Faire peur et éduquer ») : 163-186. <https://doi.org/10.3406/mar.1998.1672>.

Berlioz, Jacques. 1991. « Un petit chaperon rouge médiéval ? La petite fille épargnée par les loups dans la *Fecunda ratis* d'Egbert de Liège (début du XI^e siècle) ». *Merveilles et contes* 5 (2) : 246-263. <http://www.jstor.org/stable/41390297>.

Berlioz, Jacques. 2005. « Repérer les récits exemplaires. Des répertoires imprimés à une banque de données informatisée ». *Les Cahiers du Centre de recherches historiques* 35. <https://doi.org/10.4000/ccrh.3011>.

Berlioz, Jacques. 2007. « Des récits pour convaincre. La place du discours théologique et moral dans l'indexation des *exempla* médiévaux ». Dans *Classer les récits : théories et pratiques (actes du colloque international, fondation Singer-Polignac, Paris, 24-26 mai 2007)*, édité par Aboubakr Chraïbi, 297-312. Paris : L'Harmattan.

Berlioz, Jacques. 2016. « Un accès premier au patrimoine oral du Moyen Âge : Thesaurus des *exempla* médiévaux (ThEMA) ». Dans *Le Patrimoine oral : ancrage, transmission et édition dans l'espace gallo-romain*, édité par Aurélie Reusser-Elzingre et Federica Diémoz, 117-134. Berne : Peter Lang.

Berlioz, Jacques et Marie-Anne Polo de Beaulieu. 1992. *Les Exempla médiévaux. Introduction à la recherche, suivie des tables critiques de l'Index exemplorum de Frederic C. Tubach*. Carcassonne : GARAE-Hésiode.

Berlioz, Jacques et Marie-Anne Polo de Beaulieu. 2002. « Entre lieux communs et vie quotidienne : le petit peuple dans les recueils d'*exempla* des XIII^e et XIV^e siècles ». Dans *Le Petit Peuple dans l'Occident médiéval : terminologies, perceptions, réalités*, édité par Pierre Boglioli, Robert Delort et Claude Gauvard, 41-66. Paris : Éditions de la Sorbonne. <https://doi.org/10.4000/books.porsorbonne.13899>.

Boureau, Alain. 1984. « Towards a Microcomputerized General Index of Narratives Structures in the *Exempla* [Abstract IX] ». Dans *Medieval Sermon Studies Symposium IV*, Oxford, 12-13 juillet. Coventry : University of Warwick.

Bremond, Claude, Jacques Le Goff et Jean-Claude Schmitt. 1996 [1982]. *L'Exemplum*. Turnhout : Brepols.

Burghart, Marjorie. 2005. « L'informatisation du ThEMA. Indépendance, liberté et multilinguisme ». *Les Cahiers du Centre de recherches historiques* 35 : 79-80. <https://doi.org/10.4000/ccrh.3025>.

Collomb, Pascal. 2007. « Élaboration et utilisation du *Thesaurus Exemplorum Medii Aevi* ». Dans *Mémoire des textes, textes de mémoire*, édité par Claude Cazalé-Bérard, 51-61. Paris : Presses universitaires de Paris 10.

McGuire, Brian Patrick. 2010. *Friendship and Community. The Monastic Experience, 350-1250*. Ithaca : Cornell University Press.

Polo de Beaulieu, Marie-Anne, éd. 1991. *La Scala coeli de Jean Gobi*. Paris : CNRS Éditions.

Polo de Beaulieu, Marie-Anne. 1998. « Le "lundi des Trépassés". Création, diffusion et réception d'un rituel ». *Annales* 53 (6) : 1191-1217. <https://doi.org/10.3406/ahess.1998.279720>.

Polo de Beaulieu, Marie-Anne et Pascal Collomb. 2009. « The *Thesaurus Exemplorum Medii Aevi* (ThEMA) : An Introduction ». *Medieval Sermon Studies* 53 (1) : 45-52. <https://doi.org/10.1179/136606909X12458556541293>.

Polo de Beaulieu, Marie-Anne et Jacques Berlioz. 2005. « Introduction. Le ThEMA : l'avenir d'une base de données sur Internet ». *Les Cahiers du Centre de recherches historiques* 35. <https://doi.org/10.4000/ccrh.3008>.

Polo de Beaulieu, Marie-Anne et Pascal Collomb. 2005. « La grille de saisie de ThEMA : genèse et évolutions ». *Les Cahiers du Centre de recherches historiques* 35. <https://doi.org/10.4000/ccrh.3012>.

Thompson, Stith. 1975 [1932]. *Motif-Index of Folk-Literature. A Classification of Narrative Elements in Folktales, Ballads, Myths, Fables, Mediaeval Romances, Exempla, Fabliaux, Jest-books and Local Legends. Revised and Enlarged Edition.* Bloomington et Londres : Indiana University Press.

Tubach, Frederic C. 1969. *Index exemplorum. A Handbook of Medieval Religious Tales.* Helsinki : Suomalainen tiedeakatemia.

Welter, Jean-Thiébaud. 1973 [1927]. *L'Exemplum dans la littérature religieuse et didactique du Moyen Âge.* Genève : Slatkine Reprints.

Notes

1 Voir la liste des recueils indexés sur <http://thema.huma-num.fr/collections/list/> et les répartitions des recueils sur <http://thema.huma-num.fr/about/>.

2 Accès à l'ensemble des données : <http://thema.huma-num.fr/collections/list> ; à partir du printemps 2022, les fichiers TEI seront également archivés sur Nakala, un service d'Huma-Num (voir ci-dessous).

3 <https://tei-c.org>.

4 <http://thema.huma-num.fr/download/>.

5 <http://viaf.org>.

6 <https://www.zotero.org/groups/2304628/bibliex/>.

7 <https://www.huma-num.fr>.

8 <https://www.go-fair.org/fair-principles/>.

9 L'intégration de ThEMA dans le service de Nakala était en cours au moment de la rédaction de cet article.

10 <http://thema.huma-num.fr/sources/list/>.

Auteurs

Jean-Paul Rehr

UMR 5648 CIHAM, université Lumière Lyon 2, Lyon, France

Jean-Paul Rehr, auparavant cadre en informatique en Amérique du Nord, achève actuellement, à l'université Lumière Lyon 2, un doctorat en histoire médiévale sur les débuts de l'inquisition.

ORCID 0000-0002-9789-5669

jean-paul.rehr@univ-lyon2.fr

Marie-Anne Polo de Beaulieu

UMR 8558 CRH, CNRS et École des hautes études en sciences sociales, Paris, France

Marie-Anne Polo de Beaulieu, docteure en histoire médiévale et habilitée à diriger des recherches, est directrice de recherche au CNRS, spécialisée dans la didactique religieuse durant le Moyen Âge central européen.

ORCID 0000-0002-7617-4870

marie-anne.polo@ehess.fr

Droits d'auteur



Les contenus de la revue *Humanités numériques* sont mis à disposition selon les termes de la [Licence Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/).