



HAL
open science

PolimaWiki : un site contributif pour l'étude du pouvoir des listes au Moyen Âge

Daide Gherdevich

► To cite this version:

Daide Gherdevich. PolimaWiki : un site contributif pour l'étude du pouvoir des listes au Moyen Âge. Médiévales, 2017, Le texte à l'épreuve du numérique, 73, pp.149-167. 10.4000/medievales.8221 . hal-02134060v1

HAL Id: hal-02134060

<https://hal.science/hal-02134060v1>

Submitted on 20 May 2019 (v1), last revised 12 Nov 2020 (v2)

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.

PolimaWiki : un site contributif pour l'étude du pouvoir des listes au Moyen Âge

Polimawiki : a Semantic Wiki for Collaborative Studies about the Powers of the List in the Middle Ages

Davide Gherdevich
Université Paris I (Lamop)

PolimaWiki : un site contributif pour l'étude du pouvoir des listes au Moyen Âge
L'un des premiers objectifs du projet ANR Polima (POUvoir des Listes au Moyen Âge), débuté en 2015 à l'université de Versailles-Saint-Quentin, a été de créer un moyen de partager entre chercheurs reproductions d'objets et de documents comme éditions de textes médiévaux contenant des listes. Grâce au logiciel *Mediawiki*, et à l'extension *Semantic Mediawiki*, les données sont intégrées à un site internet collaboratif, qui en facilitera le partage et l'échange par la communauté des chercheurs.

Mots-clés : bases de données, humanités numériques, listes, PolimaWiki.

Polimawiki: a Semantic Wiki for Collaborative Studies about the Powers of the List in the Middle Ages

One of the first objectives of the ANR project Polima (Power of the lists in the Middle Ages), begun in 2015 at the University of Versailles-Saint-Quentin, was to create a means of sharing reproductions of objects and documents, among researchers as well as editions of medieval texts containing lists. Using Mediawiki software and the Semantic Mediawiki extension, the data has been integrated into a collaborative website, which facilitates sharing and exchange by the research community.

Keywords: Digital Bases, Digital Humanities, List, PolimaWiki.

Le projet ANR POLIMA¹, qui a débuté en 2015 au laboratoire DYPAC de l'université de Versailles–Saint-Quentin-en-Yvelines sous la direction de Pierre Chastang, en partenariat avec le Lamop (Université Paris 1 Panthéon-Sorbonne), sous la responsabilité scientifique et technique de Laurent Feller, est un projet interdisciplinaire (Lettres, Histoire, Linguistique, Anthropologie et Sciences Cognitives) consacré à l'étude et à la caractérisation du pouvoir des listes au Moyen Âge. À partir d'une analyse de la production, des usages et de la transmission des listes au Moyen Âge, il a pour objectif, d'une part, de contribuer à l'étude de la culture médiévale de l'écrit, et d'autre part, d'aborder de manière comparative le rapport qui s'établit dans la liste entre des énoncés linguistiques, des procédés écrits de balisage textuel et des systèmes de connaissance du monde et de contrôle des biens et des hommes.

La liste est une forme d'écriture commune à de très nombreuses productions textuelles du Moyen Âge dont la présence dans les textes, notamment « documentaires », s'accroît très fortement avec le tournant pragmatique² que connaissent les sociétés occidentales au XIII^e siècle³. Elle est fréquemment utilisée par les scripteurs et constitue, dans certains cas, un type textuel en soi, employé dans des contextes très variés. Il peut s'agir de structurer l'information à propos d'une action, d'un sujet ou d'un domaine du savoir (concordances, listes d'auteurs, souscriptions, index...), comme d'énumérer des propriétés et des parties d'objets dotés d'une consistance matérielle (inventaires, catalogues, éloges urbains) ou enfin de produire des taxinomies territoriales et fiscales.

Prenant acte des importantes variations affectant pour un même texte la forme de connexion des éléments rassemblés, nous avons retenu, comme cadre de travail, une définition de la liste de nature syntaxique. Tout texte construit selon une forme

1 POUvoirs de la LIste au Moyen Âge, ANR-14-CE24-0005.

2 H. KELLER, « Pragmatische Schriftlichkeit im Mittelalter. Erscheinungsformen und Entwicklungsstufen. Einführung zum Kolloquium », dans H. KELLER, K. GRUBMÜLLER et N. STAUBACH éd., *Pragmatische Schriftlichkeit im Mittelalter. Erscheinungsformen und Entwicklungsstufen*, Munich, 1992, p. 1-7.

3 É. ANHEIM, P. CHASTANG, « Introduction. Les pratiques de l'écrit dans les sociétés médiévales (VI^e -XIII^e siècle) », *Médiévales* [en ligne], 56 (2009), mis en ligne le 21 septembre 2011 : <http://medievales.revues.org/5515>.

parataxique⁴ peut ainsi entrer dans le corpus, ce qui conduit à englober ce que le vocabulaire courant désigne comme des énumérations, listes, catalogues, inventaires, dénombrements, etc. La notion de « pouvoir des listes », quant à elle, cherche à tracer une voie qui s'efforce de « saisir la manière dont l'écriture est toujours prise dans des contextes culturels et des rapports de pouvoir particuliers », tout en visant à une « généralisation sur ses effets, qu'ils soient cognitifs ou sociaux⁵ ».

Un des principaux objectifs du projet POLIMA a été de trouver les solutions les plus adaptées aux besoins des chercheurs, qui souhaitent disposer d'un espace de travail collaboratif dédié aux listes pour pouvoir y déposer des documents, avoir par ce biais des échanges scientifiques commodes entre spécialistes de différents domaines et contribuer ainsi à l'élaboration d'un savoir scientifique. Pour ce faire sont utilisées des technologies et des méthodologies qui permettent, entre autres, de travailler sur les sources et les données historiques dans le cadre de ce que l'on peut appeler les humanités numériques. Nous avons donc décidé de créer une plateforme en ligne⁶ pour permettre aux chercheurs, d'un côté de télécharger et de partager entre eux des reproductions de manuscrits et des documents relatifs aux listes, de l'autre d'avoir un site où publier toutes les informations et tous les documents produits pendant le projet (comme par exemple les textes de cadrage issus des discussions entre les différents membres de l'équipe, les vidéos des contributions de chercheurs de différentes disciplines lors des *workshops* organisés dans le cadre du projet et/ou tout autre matériau scientifique produit durant celui-ci). Dans la plupart des cas, le volant numérique des projets en sciences humaines et sociales (SHS) est développé seulement durant les dernières années du projet, bien après la conception scientifique, voire la réalisation, du travail. L'interface web qu'est le site ne propose alors que des textes programmatiques et il faut attendre, parfois en vain, l'exposé de résultats. Ainsi n'est-il pas possible de créer une communauté d'utilisateurs « actifs », c'est-à-dire voyant ou donnant à voir la progression du travail de recherche : recueil des matériaux, analyses intermédiaires, documents de travail, etc. C'est justement pour cette dernière raison – créer une petite

4 Juxtaposition de propositions, de groupes de mots ou de mots sans expliciter les rapports syntaxiques de subordination ou de coordination qu'ils entretiennent entre eux.

5 B. FRAENKEL, M. AÏSSATOU, « Introduction. Les *New Literacy studies*, jalons historiques et perspectives actuelles », *Langage et société*, 133/3 (2010), p. 7-24 (p. 13).

6 < <http://polima.huma-num.fr/>>.

mais solide communauté d'utilisateurs – que la technologie Open Source de type wiki, permettant l'interopérabilité des données, a été choisie pour créer le site, et ceci dès la première année du projet. Conscients des contraintes temporelles du financement et de l'organisation de la recherche par projet d'une durée de quatre ans, dans le cadre d'une ANR, nous avons également songé à l'après, à la pérennisation du travail et plus encore à sa poursuite en hébergeant le site, non pas dans l'université, mais au sein d'une structure publique nationale, précisément HUMA-NUM⁷. Avant d'exposer les choix que nous avons faits pour cette plate-forme, il nous a semblé important de les situer dans le contexte actuel du développement des Humanités Numériques.

Le contexte des humanités numériques

L'apport de ce qu'on appelle alors « l'informatique » aux travaux des sciences humaines et sociales commence dès les années 1960 avec la linguistique, pour s'étendre ensuite aux autres disciplines, telles que l'histoire et l'archéologie, à partir des années 1980. En France, les premiers travaux d'informatique appliquée à l'histoire médiévale, par exemple, sont nés dans les années 1980, sous l'impulsion notamment de Jean-Philippe Genet⁸ et des contributeurs des revues *Le médiéviste et l'ordinateur* et *Histoire & Mesure* ; ils se poursuivent au cours des années 2000 avec le programme ANR ATHIS (Ateliers Histoire et Informatique)⁹. Dans les années 1990, c'est surtout en Angleterre et aux États-Unis que, grâce au réseau internet, on assiste au développement de formes abouties des *digital humanities*, à savoir des sites, des bases de données en ligne, des systèmes d'information géographique (SIG), des projets de numérisation et d'accès en ligne à des documents manuscrits, etc. ; les communautés scientifiques, nées de ce mouvement, commencent à être plus visibles grâce aux rencontres annuelles dont le thème est précisément les *digital humanities* : le premier colloque de l'*Alliance of Digital Humanities Organizations* fut

⁷ <<http://www.huma-num.fr/>>.

⁸ J.-P. GENET, « Histoire, Informatique, Mesure », *Histoire & Mesure*, 1 (1986), p. 7-18.

⁹ J.-P. GENET et A. ZORZI, éd., *Les Historiens et l'informatique : un métier à réinventer*, Rome, 2011 (École française de Rome, 444).

organisé à l'université de Siegen¹⁰, en Allemagne, en 1990 ; le prochain colloque se tiendra à Montréal, au Canada, en 2017.

En France, les humanités numériques ont commencé à se structurer du point de vue scientifique surtout au début des années 2000, grâce à un certain nombre de projets, parmi lesquels le TGE ADONIS du CNRS¹¹, né en 2007 pour créer une infrastructure numérique (avec la plateforme ISIDORE¹²) d'accès unifié aux données et documents des Sciences humaines et sociales. En 2013, ADONIS a été fusionné avec CORPUS-IR¹³, une plateforme de coopération permettant l'accès aux principaux ensembles documentaires (images, sons, textes) produits en priorité dans le cadre de la recherche en SHS ; de l'union de ces deux grandes plateformes est née la très grande infrastructure de recherche (TGIR) HUMA-NUM. En 2010 s'est tenue à Paris la première grande rencontre avec les différents acteurs et utilisateurs des humanités numériques en SHS ; à l'issue de ce colloque, qui a pris le nom de That CAMP 2010¹⁴, le premier *Manifeste des Digital Humanities* a vu le jour¹⁵.

Malgré ce début de structuration institutionnelle, peut-on toutefois définir aisément les humanités numériques ? Encore aujourd'hui, nous sommes au regret d'observer qu'il n'en existe pas de définition satisfaisante, ni unifiée pour tous. Dans le *Manifeste*, déjà cité, « les humanités numériques désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales¹⁶ ». Frédéric Kaplan¹⁷, quand il parle sur son blog des humanités numériques, affirme qu'il s'agit

10 <Adho.org>.

11 Très grand équipement (TGE) ADONIS du Centre National de la Recherche Scientifique.

12 <https://www.rechercheisidore.fr/>

13 Coopération des Opérateurs de Recherche Pour un Usage des Sources numériques en Sciences Humaines et Sociales.

14 <<https://tcp.hypotheses.org/category/thatcamp-paris-2010>>.

15 La version originale du *Manifeste* est en ligne (<<https://tcp.hypotheses.org/318>>). Un autre manifeste a été écrit par Jeffrey Schnapp, en 2009, *The Digital Humanities Manifesto 2.0*, téléchargeable en ligne (<http://jeffreyschnapp.com/wp-content/uploads/2011/10/Manifesto_V2.pdf>).

16 *Manifeste des Digital humanities*, Conference THATCamp des 18 et 19 mai 2010, Paris, <<https://tcp.hypotheses.org/318>>.

17 Frédéric Kaplan est professeur en Humanités Numériques à l'École Polytechnique de Lausanne et idéateur du projet « Venice Time Machine ».

« d'appliquer les savoir-faire des sciences de l'information aux questions de sciences humaines et sociales¹⁸ ». Si on regarde dans le monde anglo-saxon, Jeffrey Schnapp affirme à ce sujet :

Les humanités numériques se réfèrent à de nouveaux modes d'études et d'unités institutionnelles pour la recherche, l'enseignement et la publication de travaux collaboratifs, transdisciplinaires et informatisés. Les sciences humaines numériques, plus qu'un domaine unifié, sont une série de pratiques convergentes qui explorent un univers dans lequel l'impression n'est plus le principal moyen pour produire et diffuser la connaissance. Les outils, les techniques et les médias numériques ont élargi les concepts traditionnels de connaissances dans les arts, les sciences humaines et les sciences sociales, mais les humanités numériques ne sont pas uniquement « l'utilisation » du numérique dans les sciences humaines et ne sont pas les humanités classiquement définies puisqu'elles font valoir une reconfiguration des pratiques traditionnelles. Les sciences humaines numériques sont plutôt définies par les opportunités et les défis qui découlent de la conjonction du terme numérique avec le terme humanités pour former un nouveau collectif singulier¹⁹.

Le projet qui nous intéresse ici s'inscrit, même si les contours du domaine sont encore flous, comme une réalisation des humanités numériques. Dans le domaine de l'histoire médiévale, nous pouvons ainsi donner plusieurs exemples de projets qui ont utilisé les instruments du numérique pour l'étude de

18 <<https://fkaplan.wordpress.com/tag/humanites-digitales/>>.

19 « Digital Humanities refers to new modes of scholarship and institutional units for collaborative, transdisciplinary, and computationally engaged research, teaching, and publication. Digital Humanities is less a unified field than an array of convergent practices that explore a universe in which print is no longer the primary medium in which knowledge is produced and disseminated. Digital tools, techniques, and media have expanded traditional concepts of knowledge in the arts, humanities and social sciences, but Digital Humanities is not solely “about” the digital (in the sense of limiting its scope to the study of digital culture). Nor is Digital Humanities only “about” the humanities as traditionally understood since it argues for a remapping of traditional practices. Rather, Digital Humanities is defined by the opportunities and challenges that arise from the conjunction of the term digital with the term humanities to form a new collective singular » (J. SCHNAPP, dans A. BURDICK, J. DRUCKER, P. LUNENFELD, T. PRESNER et J. SCHNAPP éd, *Digital Humanities*, Boston, 2012, p. 121-136, ici p. 142).

manuscrits. En particulier, nous pouvons citer ceux qui ont été utilisés comme référence pour la création de PolimaWiki :

– le projet CBMA (*Chartae Burgundiae Medii Aevi*) d’Eliana Magnani²⁰, qui a mis à disposition des chercheurs une plateforme documentaire de consultation et une base de données accessible sur internet et téléchargeable sous différents formats, avec une partie dédiée à la cartographie ;

– le projet ANR Thalamus²¹, basé sur le standard XML/TEI²², qui a eu comme résultat l’édition électronique du manuscrit du *Petit Thalamus* de Montpellier, mise à disposition de tous les chercheurs au fur et à mesure de l’avancée du projet ;

– le projet SCRIPTA²³, une base de données en ligne, pour le moment seulement textuelle, mais qui compte à peu près 8 500 actes normands médiévaux. L’intention des réalisateurs dans les prochaines années est d’élargir la base à d’autres corpus, y compris des actes inédits et des documents publiés en XML/TEI. On peut aussi citer deux autres projets qui, même s’ils ne sont pas dans le domaine de l’histoire médiévale, nous ont permis de voir les potentialités d’un logiciel comme MediaWiki : le Wiki réalisé par le musée Edvard Munch, dédié aux lettres et à la documentation du peintre²⁴, et le projet Hyperlistes de Madeleine Jeay, un corpus de poèmes du XIII^e au XVI^e siècle, qui n’est plus accessible en ligne mais dont le corpus est en cours d’intégration dans PolimaWiki.

Les choix techniques

La réalisation du site et de la base de données est strictement liée aux demandes des chercheurs impliqués dans le projet qui voulaient un outil au fonctionnement similaire de celui d’une base de données traditionnelle, avec toutefois les commodités d’un outil du type *wiki* conçu pour favoriser la collaboration entre les individus et la mise en relation de plusieurs documents. Pour exemple, Pierre Chastang avec les listes contenues dans le

20 <Erreur ! Référence de lien hypertexte non valide.>

21 <<http://thalamus.huma-num.fr/index.html>>.

22 XML, Extensible Markup Language, en français « langage de balisage extensible » ; TEI, Text Encoding Initiative, en français « initiative pour l’encodage du texte ».

23 *Scripta*, site de recherche informatique et de publication des textes anciens du CRHAM de l’université de Caen Normandie, <<http://www.unicaen.fr/scripta/cdc.html>>.

24

<<http://www.emunch.no/wiki/index.php/Manuskripter#Fransk>>.

« Thalamus des ouvriers de la commune clôture »²⁵, ou Laurent Feller et Nicolas Morel avec la liste des esclaves du « Polyptyque d'Irminon »²⁶, ou Madeleine Jeay qui a mis en ligne 124 listes littéraires²⁷. Nous avons donc essayé de trouver la solution technique la plus adaptée afin de répondre à ces besoins et de faciliter le travail de collaboration, pendant le projet et au-delà, car, comme nous l'avons déjà évoqué, l'autre facteur que nous avons dû prendre en considération est le temps : nous n'avons disposé que d'une année pour développer le site internet et la base de données afférente, et nous avons donc dû chercher un logiciel facile à prendre en main rapidement pour des utilisateurs à l'aise avec les outils de la bureautique, mais peu habitués en revanche à des applications dédiées à la publication de données sur internet. Il fallait trouver un logiciel déjà prêt à l'usage, nécessitant seulement quelques ajustements ou de petits développements.

Une des premières questions que nous nous sommes posée au niveau technique, avant même de réfléchir au choix du logiciel, a été de savoir quel type de philosophie de développement utiliser : à code ouvert (*Open Source*) ou sous licence commerciale. Finalement l'équipe a choisi d'opter pour le logiciel libre, notamment parce qu'il permettait de bénéficier d'un réseau de développeurs et d'utilisateurs passionnés qui pouvaient aider à résoudre les problèmes et fournir des conseils lors de la phase de réalisation. En outre, le fait que le logiciel libre n'entre pas dans la logique commerciale des entreprises privées permet d'assurer, entre autres, une « pérennité » des données²⁸, importante quand il existe, comme dans le cas qui nous intéresse, une grande communauté d'utilisateurs. Pour PolimaWIKI, il s'est agi de combiner les intérêts de deux communautés : celle des chercheurs et celle des développeurs, en donnant des exemples convaincants des usages qui peuvent être faits de ce type de rencontre.

Nous avons envisagé plusieurs solutions techniques : la première aurait été de créer une base de données classique en utilisant

25 <http://polima.huma-num.fr/wiki/List:15>

26 <http://polima.huma-num.fr/wiki/List:161>

27 Une liste des chercheurs impliqués dans le projet et tous les documents qui ont été déjà insérés dans le wiki sont visibles sur le site <http://polima.huma-num.fr/>.

28 Mais peut-on parler de pérennité pour les logiciels et les standards informatiques, étant donnée la vitesse du développement des solutions techniques ?

comme logiciel MySQL²⁹, et, dans le même temps, de développer une interface graphique, un formulaire et des modules de recherche en PHP³⁰ et HTML (nous aurions pu aussi utiliser d'autres langages multiplateformes comme par exemple Java). Cette solution est la plus utilisée pour mettre des bases des données en ligne, mais elle présente deux inconvénients : le premier est que les bases de données classiques sont souvent fermées et laissent très peu de liberté au contributeur pour leur élaboration. Même si cette rigidité représente l'un de leurs principaux points de force, elle peut toutefois créer quelques difficultés lors de la compilation. À l'inverse, mais tout aussi embarrassant, une base de données très ouverte peut causer plusieurs problèmes, surtout dans la réalisation des modules de recherche. Un autre point a joué en défaveur de cette option : l'interface et la structure du site devaient alors être entièrement développées et, comme nous l'avons dit, le temps dédié au développement était d'un an seulement ; après cette période, nous n'avions pas la certitude de disposer d'un ingénieur pour travailler sur la structure de la base et sur l'interface. Une autre solution possible, pour développer plus rapidement le site, pouvait être l'utilisation d'un Système de gestion de contenu ou CMS (*Content Management System*). Parmi les nombreux logiciels CMS, Wordpress, Joomla ou Drupal, par exemple, sont des systèmes plutôt faciles à utiliser et plutôt adaptables aux besoins de l'équipe, mais malheureusement les modules d'extension dédiés à la création de formulaires complexes, comme les systèmes de recherche et d'affichage, y sont très peu développés et plutôt limités.

Le logiciel qui s'est finalement révélé être le plus adapté et que nous avons décidé d'utiliser pour la création de notre site internet est le logiciel MediaWiki³¹, CMS de la fameuse *Wikipédia*, l'encyclopédie contributive consultable en ligne. Avant d'arrêter notre choix sur MediaWiki, nous avons essayé deux différentes variantes, XWIKI et Tiki WIKI, de la même « famille » que MediaWiki, mais dont les communautés sont

29 MySQL est un système de gestion de bases de données relationnelles distribué sous une double licence libre et propriétaire. Il fait partie des logiciels de gestion de base de données les plus utilisés au monde (<<https://www.mysql.fr/>>).

30 PHP (officiellement, ce sigle est un acronyme récursif pour PHP Hypertext Preprocessor) est un langage, spécialement conçu pour le développement d'applications web. Il peut être intégré facilement au HTML (<<http://php.net/manual/fr/intro-what-is.php>>).

31 <<https://www.mediawiki.org/wiki/MediaWiki>>.

moins importantes et, dans le cas de XWIKI, avec des outils payants et liés à une entreprise commerciale.

Le logiciel MediaWiki est un logiciel libre distribué selon les termes de la GPL³², développé par la Wikimedia Foundation et conçu pour répondre aux besoins de *Wikipédia*. Depuis 2008, MediaWiki est également utilisé par des entreprises comme solution de gestion des connaissances et comme système de gestion de contenu. Le logiciel est écrit en PHP et peut aussi bien fonctionner avec le système de gestion de base de données MySQL que PostgreSQL. C'est un logiciel avec de nombreuses fonctionnalités et des outils pour les sites à vocation collaborative³³. Il répond donc parfaitement à l'objectif du projet POLIMA qui entend non seulement créer un site où stocker les données, mais aussi être un espace de participation et de collaboration active entre les chercheurs (par exemple, en apportant des corrections aux transcriptions, en faisant des liens entre les listes, ou encore en ouvrant des discussions autour des documents).

Grâce aux nombreux modules d'extension, nous avons pu adapter le logiciel à nos besoins. Parmi tous les modules existants, nous en avons utilisé un en particulier : l'extension Semantic MediaWiki et toutes ses sous-extensions (fig. 1).

Fig. 1. Le site Polima.hum-num.fr

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Semantic MediaWiki et Page Form

Le mot « sémantique » se réfère ici à la possibilité de représenter de manière formelle et standardisée la nature et la valeur des contenus des ressources web, ainsi que les relations entre eux³⁴. Semantic MediaWiki est une extension qui permet d'ajouter des

32 General Public License.

33 J. J. JUNG, « Computational Reputation Model Based on Selecting Consensus Choices : an Empirical Study on Semantic Wiki Platform », *Expert System with Applications*, 39-10 (2012), p. 9002-9007.

34 M. COCCOLI *et al.*, « I Semantic Wiki come strumento per la progettazione collaborativa di contenuti didattici », *Journal of e-learning and Knowledge Society - IT*, 8-2 (2012), p. 115-224 (p. 116).

annotations sémantiques aux pages d'un wiki. Les annotations ajoutées peuvent ensuite être utilisées pour réaliser des recherches sémantiques, pour agréger des pages entre elles et structurer leur contenu de manières différentes (comme, par exemple, sur une carte géographique, un calendrier, un graphe), ou pour exporter les données afin qu'elles puissent être utilisées par des applications tierces. En résumé, Semantic MediaWiki unit la logique des wikis, des bases de données et du web sémantique et permet de créer des applications plus structurées³⁵.

Semantic MediaWiki nous permet « [de] rechercher, organiser, baliser, naviguer, évaluer et partager le contenu d'un wiki. Alors que les wikis traditionnels ne peuvent contenir que du texte (ne pouvant être ni compris, ni évalué par les ordinateurs), Semantic MediaWiki ajoute des annotations sémantiques permettant à un wiki de fonctionner comme une base de données collaborative³⁶. »

Cet outil nous permet de rechercher presque tous les mots contenus dans les documents et de créer aussi des pages où les résultats des recherches sont organisés sous forme de tableaux, de listes bibliographiques ou de n'importe quelle autre forme configurable par des modèles, qui se mettront ensuite à jour automatiquement. Par exemple, si nous ajoutons ou changeons la valeur d'un champ dans notre modèle de liste, il ne sera pas nécessaire d'aller chercher pour corriger cette valeur dans tous les tableaux associés à la liste, le changement aura été simultané.

Le cœur du dispositif Semantic MediaWiki est l'attribut ou la propriété, que nous pouvons définir comme « une catégorie d'information » qui décrit formellement le contenu d'une page. Un attribut est composé d'un nom (identificateur) et d'une valeur qui peut être par exemple du type « nombre », « texte », « URL », « date », « page ». Lorsqu'on insère une propriété plus une valeur dans une page, on définit implicitement un « triplet », qui est la base du concept du web sémantique. Le « triplet » à niveau sémantique est composé d'un « sujet », d'un « prédicat » et d'un « objet » ; dans Semantic MediaWiki, le « sujet » est la page de notre wiki, le « prédicat » est la propriété et l'« objet » est la valeur de la propriété.

Pour remplir la base de données et créer les pages de notre wiki, nous avons utilisé une autre extension, toujours liée à Semantic

35 M. KRÖTZSCH et alii, « Semantic Wikipedia », *Journal of Web Semantics*, 5 (2007), p. 251-216.

36 https://www.semantic-mediawiki.org/wiki/Help:_Pr%C3%A9sentation_de_Semantic_MediaWiki.

MediaWiki, qui s'appelle Page Form. Cette extension permet de créer des formulaires, grâce auxquels nous pouvons ajouter et éditer les pages ; pour faire cela, on utilise des champs qui peuvent être remplis avec les classiques boîtes combo³⁷, les *checkboxes*, les radio-boutons, ou encore avec les zones de texte d'auto-complétion, pour permettre de restreindre ou de proposer des valeurs aux utilisateurs.

La structure du formulaire

Grâce aux extensions Page Form et Semantic MediaWiki, nous avons commencé à réaliser notre site internet et à structurer notre formulaire. Celui-ci a été réalisé avec la collaboration de la plupart des chercheurs et ingénieurs engagés sur le projet POLIMA.

L'utilisateur peut accéder au formulaire dans la page d'accueil en utilisant le bouton « déposer » (fig. 2). Chaque fois qu'il ouvre un formulaire, le logiciel donne un nom univoque à la liste qui va être ainsi créée. Dans la nouvelle page qui s'affiche, le contributeur est face à la première page du formulaire et aux onglets qui permettent d'accéder aux autres parties de ce dernier. Ce formulaire est structuré en quatre parties :

- identifier ;
- décrire ;
- déposer ;
- documenter.

Fig. 2. Page d'accueil et bouton « déposer la liste »

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Identifier. Cette première partie du formulaire est dédiée aux informations sur le texte ou le manuscrit : le nom de l'auteur (si nous le connaissons), le titre de l'œuvre qui contient la liste ou, plus rarement, le nom de la liste elle-même. Pour créer une homogénéité graphique des noms comme des titres, on a placé à côté des champs un lien pour les notices d'autorité de la BnF³⁸. Après cela, le champ « incipit » permet d'insérer la première

37 Une boîte combinée (en anglais combo-box ou combobox) est un élément d'interface graphique qui réunit une zone de texte et une liste déroulante.

38

<http://www.bnf.fr/fr/professionnels/donnees_autorites.html>.

ligne du document. Celui dénommé « commentaire libre », qui est le seul présent dans les quatre parties du formulaire, et dans lequel les chercheurs peuvent donner des indications plus ponctuelles (commentaires, précisions factuelles, etc.) qui ne se trouvent pas dans le champ du formulaire (fig. 3).

Fig. 3. Identifier

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Décrire. Cette partie est dédiée à la description de la liste et est composée de plusieurs champs : le premier permet de comprendre la structure de la liste, c'est-à-dire d'identifier si elle est visuellement continue ou si elle procède par retours à la ligne. Le deuxième se réfère à la période de production de la liste ; nous avons choisi de faire plusieurs propositions de dates et, ainsi, de pouvoir effectuer des choix multiples (par exemple la date du texte et la date du témoin qui le conserve). Le troisième nous permet de sélectionner la langue, et le quatrième le lieu de production ; dans ce dernier cas, pour éviter des problèmes liés à la localisation du lieu, nous avons prévu deux solutions possibles :

- le positionnement du lieu de production de la liste directement sur une carte géographique grâce à l'utilisation de google maps ;
- l'option de choisir le pays et la région à travers un menu déroulant, en utilisant comme référence les limites des pays et des régions actuels. Nous n'avons pas choisi le découpage géographique de l'époque médiévale pour des raisons évidentes, du fait de la complexité de l'évolution des limites entre les différentes entités territoriales.

Les derniers champs sont relatifs aux typologies des listes : comme base de référence nous avons utilisé *La Typologie des sources du Moyen Âge occidental*³⁹, avec des élaborations et des intégrations effectuées par les historiens du projet afin de rendre la liste la plus exhaustive possible (fig. 4).

Fig. 4. Décrire

© Crédits ? source ?

39 L. GENICOT, « La typologie des sources du Moyen Âge occidental », *Annales. Économies, Sociétés, Civilisations*, 6 (1972), p. 1257-1263.

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Déposer. Cette partie du formulaire est composée de deux parties, dédiées aux informations complémentaires liées à la liste et, quand cela est possible, à l'insertion des images⁴⁰ et d'une transcription de la liste ou du renvoi à l'édition du texte. Pour faciliter le travail des chercheurs, nous avons décidé de créer deux formulaires différents : un pour les listes décrites à partir d'éditions de textes modernes, et un autre pour celles contenues dans des témoins manuscrits. Dans les deux cas, il est possible de télécharger les photos de la liste ; s'il s'agit d'une édition moderne, on peut choisir de remplacer celle-ci par un lien hypertexte qui permet d'accéder au site où le document est visible (c'est le cas par exemple des documents numérisés dans Gallica⁴¹). En ce qui concerne les images, plusieurs formats sont acceptés : jpg, pdf, png ; le système reconnaît également les métadonnées liées à la photo : si par erreur le contributeur télécharge la même photo, un message d'erreur s'affiche pour éviter de se retrouver avec plusieurs copies inutiles du même document.

Les deux formulaires comptent plusieurs champs identiques, comme les champs « ville », « institution », « cote » et « transcription ». C'est seulement pour les listes décrites à partir de leur forme manuscrite que nous disposons de champs sur les informations relatives au support et, plus spécifiquement, au matériel sur lequel la liste a été écrite. Grâce à un menu déroulant, le chercheur peut sélectionner les différents matériaux (parchemin, papier, etc.), la forme du support (inscription, rouleau, feuille, etc.), le numéro du folio et s'il s'agit du recto ou du verso. En revanche, pour les éditions modernes, nous avons des champs spécifiques : l'éditeur scientifique, la date de publication, le titre et les pages où se trouve la liste. À noter

40 Pour la France, l'Italie et l'Europe en général, pour des publications sur internet à caractère scientifique ou pédagogique qui ne donnent pas lieu à rémunération de droits d'auteur, la licence de réutilisation des images est gratuite. Pour avoir plus d'information sur les droits d'auteur : <http://www.bnf.fr/fr/professionnels/droit_auteur.html>.

41 Dans ce cas il suffit d'utiliser le permalien, un type spécial d'URL conçu pour rester inchangé sur une très longue période ; ce permalien (pemalink en anglais) est normalement mis à disposition sur le portail où est conservée l'édition numérisée (Gallica, Archive.org, etc...).

que, grâce au webservice fourni par l'IRHT – qui nous a permis de sortir en format XML (mais aussi en JSON⁴²) la liste des villes et des institutions de la base de données Medium⁴³ –, il est possible de remplir automatiquement les champs « ville », « pays » et « institution » de notre formulaire.

Pour le moment, dans le champ « transcription », l'utilisateur peut seulement télécharger le fichier en format .txt, qui toutefois ne permet pas de conserver la mise en page originale de la liste. Enfin, un bouton positionné à la fin du formulaire donne encore la possibilité d'ajouter d'autres images et transcriptions, en évitant de remplir de nouveau les champs relatifs aux informations sur le manuscrit ou l'édition (fig. 5).

Fig. 5. Formulaire dédié aux listes des manuscrits

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Documenter. Le dernier onglet du formulaire est dédié à la bibliographie de la liste et à l'identification du contributeur. Pour la bibliographie, l'utilisateur peut ajouter des monographies, des articles et des ouvrages collectifs, qui iront alimenter la page de la bibliographie générale du projet. Dans cette dernière partie se trouve le seul champ obligatoire de tout le formulaire, dans lequel il faudra insérer le nom de la personne qui a référencé la liste (fig. 6).

Fig. 6. Documenter

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Une base de données collaborative et sémantique

Tous les documents recueillis constituent une base de données consultable en ligne qui peut être alimentée et amendée au fil de

⁴² *JavaScript Object Notation.*

⁴³ *Medium*, répertoire des manuscrits reproduits et recensés de l'IRHT (<<http://medium.irht.cnrs.fr/>>).

la progression du travail des chercheurs. En effet, le but du projet POLIMA n'est pas en soi la création d'une base de données, ou du moins cette dernière a pour finalité d'être un outil collaboratif: chaque chercheur a la possibilité d'apporter des modifications, en améliorant la transcription des textes, en ajoutant des notes ou en faisant des liens avec d'autres documents. Pour chaque liste, on peut créer une page de discussion, à laquelle tous les chercheurs seront appelés à participer, pour développer le wiki et contribuer à l'enrichissement d'un débat autour des documents.

MediaWiki permet en outre de gérer plusieurs niveaux d'accès au contenu du site. Durant le projet ANR, la possibilité d'ajouter des listes et l'accès aux parties liées aux workshops sont réservés aux chercheurs qui participent au projet, et éventuellement aux chercheurs extérieurs qui en font la demande au préalable pour obtenir l'autorisation des administrateurs du site. Lorsque le projet ANR sera achevé, il sera possible d'effectuer son inscription directement en utilisant une procédure automatisée, la modération n'intervenant qu'avant publication des contributions soumises. Précisons qu'à tout moment tout utilisateur du web peut, *via* le site en ligne, effectuer des recherches et d'exporter les métadonnées liées à la liste qui l'intéresse au format RDF.

Une fois que nous avons créé la base, Semantic MediaWiki nous permet de rechercher des informations spécifiques en utilisant plusieurs systèmes : du plus classique (à travers un simple mot) au plus avancé, comme la recherche sémantique qui autorise des recherches croisées sur plusieurs champs. Par exemple, chaque champ étant associé à une propriété, nous pouvons rechercher toutes les listes écrites en latin entre 1275 et 1300 (fig. 7) ; il suffit d'écrire dans le champ requête :
[[Langue::Latin]][[Période::1275-1300]].

Fig. 7. Recherche sémantique

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Enfin, si nous voulons voir figurer dans les résultats d'autres informations, il suffit d'ajouter les noms des propriétés dans le champ à côté, nommé « données supplémentaires à afficher » (on peut par exemple ajouter le type de liste, le support, la typologie, etc.). Le résultat donnera lieu à un tableau, qui

rassemblera toutes les informations que nous avons choisi d'afficher (fig. 8).

Fig. 8. Tableau avec les informations que nous avons choisi d'afficher

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Semantic MediaWiki permet aussi de réaliser une recherche par filtre (Semantic Drilldown), en utilisant les champs du formulaire pour filtrer les informations ; par exemple, on peut choisir comme premier filtre tous les documents en langue vernaculaire et, ensuite, affiner notre recherche en ajoutant d'autres filtres comme le période ou le pays (fig. 9).

Fig. 9. Recherche par filtre (Semantic Drilldown)

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

En outre, nous pouvons aussi effectuer des sélections à partir des métadonnées qui sont visibles à la fin de chaque liste, simplement en utilisant le symbole de la loupe, qui permet de chercher des pages similaires (fig. 10). Pour permettre l'intégration de toutes ces informations dans d'autres bases de données, nous pouvons ensuite exporter directement les données en format RDF⁴⁴ (*via* un lien sur le panneau des métadonnées) ou en format XML.

Fig. 10. Recherche de pages similaires

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Un autre élément a été développé dans le wiki, concernant le positionnement géographique des listes. Au moyen de

⁴⁴ *Resource Description Framework.*

l'extension Semantic Map⁴⁵, nous pouvons positionner chaque manuscrit dans l'espace géographique, tandis que la fonction *geotagging* permet d'effectuer de nombreuses recherches et analyses, non seulement liées au contenu du document, mais aussi à sa localisation (nous pouvons ainsi exporter les coordonnées géographiques des listes en format KML, le format des fichiers de Google Earth (fig. 11)).

Fig. 11. Localisation de liste sur Google earth

© Crédits ? source ?

IMAGE MANQUANTE : MERCI DE NOUS FOURNIR AU PLUS VITE UN FICHIER-IMAGE HAUTE-DEFINITION (= EXPLOITABLE POUR L'IMPRIMÉ)(NOIR ET BLANC OU COULEUR) laurence.hallouin@univ-paris.fr

Nous avons voulu, dans cet article, exposer principalement les fonctionnalités liées à la base de données sémantiques accessibles le site à tout chercheur. Mais le site PolimaWiki permet aussi de conserver et de consulter tous les documents et les vidéos qui ont été produits, pendant le projet POLIMA, en se transformant en un Bliki. Ce mot, qui vient de l'union de « blog » et « wiki »⁴⁶, met en évidence l'ambition même du projet, reposant sur l'idée que ce sont les chercheurs qui vont, en collaborant, créer une collection de listes ; les données ne sont pas simplement « conservées » à l'intérieur de la base, elles sont aussi partagées par la communauté pour permettre à chacun d'enrichir, avec ses expériences, la connaissance sur les listes du Moyen Âge. Les possibilités ouvertes par le wiki sémantique sont énormes ; avec un peu de pratique, chaque utilisateur pourra construire dans PolimaWiki un véritable réseau, où les différentes listes seront liées entre elles par des liens sémantiques ; il pourra créer des pages, en juxtaposant des listes de la base avec des documents issus des workshops, ou encore établir des liens avec d'autres sites afin de mettre en pratique l'interopérabilité qui est à la base du web sémantique. En ceci, POLIMA a su créer, en peu de temps, une plateforme collaborative qui s'inscrit parfaitement dans le « paysage » des humanités numériques en cours de structuration.

45 Depuis novembre 2016, l'extension s'appelle *Maps* (« <https://www.semantic-mediawiki.org/wiki/Extension:Maps> »).

46

<https://www.mediawiki.org/wiki/Manual:Using_MediaWiki_as_a_content_management_system>.

Au moment de la rédaction de cet article, nous comptons cinquante-deux utilisateurs enregistrés, dont dix-huit sont des contributeurs actifs qui ont mis en ligne 173 listes : peut-être avez-vous une liste à ajouter ?