



**HAL**  
open science

# Textes savants et pratiques numériques : le projet Poésie biblique latine antique et médiévale (PoBLAM)

Bruno Bureau, Inès Burri

## ► To cite this version:

Bruno Bureau, Inès Burri. Textes savants et pratiques numériques : le projet Poésie biblique latine antique et médiévale (PoBLAM). *Humanités numériques*, 2024, Dictionnaires numériques - varia, 10 (10), 10.4000/12ypw . hal-04810984

**HAL Id: hal-04810984**

**<https://hal.science/hal-04810984v1>**

Submitted on 17 Jan 2025

**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

## Textes savants et pratiques numériques : le projet *Poésie biblique latine antique et médiévale (PoBLAM)*

*Scholarly Texts and Digital Practices: the Poésie biblique latine antique et médiévale (PoBLAM) Project*

Bruno Bureau et Inès Burri

---



### Édition électronique

URL : <https://journals.openedition.org/revuehn/4218>

DOI : 10.4000/12ypw

ISSN : 2736-2337

### Éditeur

Humanistica

Ce document vous est fourni par INIST - Centre national de la recherche scientifique (CNRS)



### Référence électronique

Bruno Bureau et Inès Burri, « Textes savants et pratiques numériques : le projet *Poésie biblique latine antique et médiévale (PoBLAM)* », *Humanités numériques* [En ligne], 10 | 2024, mis en ligne le 01 décembre 2024, consulté le 17 janvier 2025. URL : <http://journals.openedition.org/revuehn/4218> ; DOI : <https://doi.org/10.4000/12ypw>

---



Le texte seul est utilisable sous licence CC BY 4.0. Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.



## **Textes savants et pratiques numériques : le projet *Poésie biblique latine antique et médiévale (PoBLAM)***

### ***Scholarly Texts and Digital Practices: the Poésie biblique latine antique et médiévale (PoBLAM) Project***

**Bruno Bureau et Inès Burri**

#### **Résumés**

Cet article vise à partager les réflexions qu'ont suscitées certaines composantes numériques du projet *Poésie biblique latine antique et médiévale (PoBLAM)* financé par l'Agence nationale de la recherche (ANR) et la Deutsche Forschungsgemeinschaft (DFG). Il s'agit de présenter, à partir de questions plus générales que pose l'utilisation dans ce contexte des humanités numériques, la manière dont les aspects techniques ont influencé nos plans de réalisation. Bien que tous les projets portent sur un sujet commun, la poésie biblique latine et médiévale, nous avons été contraints d'adopter des approches très diverses en raison de leurs différences de nature : ce sont en effet un hypertexte, une édition savante et une base de données. La deuxième partie de l'article offre une explication détaillée et argumentée de nos choix techniques dans le contexte précis des attentes de *PoBLAM*.

The aim of this article is to share the reflections arising from some digital components of the ANR-DFG *PoBLAM* project. Based on the more general questions raised by the use of digital humanities in this context, we present the impact of the technical aspects of the project on our implementation plans. Although all the projects deal with a common subject, Latin and medieval biblical poetry, we had to adopt very different approaches because of the differences in their nature: they are in fact a hypertext, a scholarly edition and a database. The second part of the article offers a detailed explanation of our technical choices and their justification in the precise context of *PoBLAM*'s expectations.

## Entrées d'index

MOTS-CLÉS : littérature, études anciennes et études classiques, édition savante numérique, base de données, TEI, intertextualité

KEYWORDS: literature, ancient studies and classical studies, scholarly digital edition, database, TEI, intertextuality

## Introduction

<sup>1</sup> Le projet de recherche franco-allemand *Poésie biblique latine antique et médiévale (PoBLAM<sup>1</sup>)* financé par l'Agence nationale de la recherche (ANR) et la Deutsche Forschungsgemeinschaft (DFG), est né de la volonté de réunir divers points de vue sur l'évolution de la poésie en milieu chrétien latin et grec dans l'Antiquité tardive et au Moyen Âge, ainsi que sur le rôle de la poésie dans l'évolution de la culture, tout en mettant en dialogue les pratiques française et allemande dans le domaine de la philologie et de l'interprétation des textes. À côté de colloques et de production d'ouvrages ou d'articles, un axe important est la création d'objets numériques, d'éditions numériques et de bases de données, illustrant les divers champs d'action envisagés. S'agissant d'un projet qui s'adosse à d'autres initiatives fédératives internationales sur une thématique propre comme le GIS (Groupement d'intérêt scientifique) international du Girpam (Groupe international de recherches sur la poésie de l'Antiquité tardive et du Moyen Âge<sup>2</sup>), l'important était de penser dès le début tous les objets numériques qui seraient créés dans une logique d'élargissement possible à d'autres thématiques, d'autres textes ou d'autres contextes. Il fallait également que les éléments conçus ne soient pas indépendants les uns des autres mais puissent dialoguer entre eux, soit dans la première phase actuellement en réalisation, soit dans les phases ultérieures de la recherche, les groupes constitués autour de ces projets ayant vocation à perdurer au-delà de la durée du contrat ANR-DFG.

<sup>2</sup> Les projets numériques centraux de cette action sont *Poéxégèse*, *Barbosa* et *Vestigia Iuuenci*<sup>3</sup>. Ils relèvent essentiellement de deux types : des projets orientés textes et des projets orientés données. À première vue, il n'y avait là rien de bien nouveau et il était théoriquement possible de s'appuyer sur des éléments existants. Or, au fil du projet, la complexité des demandes des partenaires, comme la nécessité de penser un modèle global réutilisable et adaptable sans trop de difficultés, a fait apparaître des zones grises, en particulier lorsqu'on se plaçait à l'intersection des projets orientés base de données et des projets orientés textes.

<sup>3</sup> Toute la question méthodologique<sup>4</sup> dans ce type d'initiative de recherche porte alors sur la manière d'envisager l'interaction entre les corpus textuels et leurs données, ou à l'inverse entre des jeux de données préalablement construits et des corpus textuels. Pour cela, il faut envisager le problème sous ses deux aspects, celui de la construction des bases de données et de leur structuration en anticipant l'exploitation des résultats dans un cadre d'annotation textuelle, et celui de l'encodage textuel, de manière à rendre possible l'importation totale ou partielle d'éléments issus de requêtes faites sur la ou les bases de données. Or, dans cette

double approche, une nouvelle logique se fait jour, celle de l'alimentation automatique de la ou des bases de données par des éléments issus de l'encodage textuel (données semi-structurées) et, à l'inverse, la projection sur l'encodage textuel de tout ou partie des données structurées dans la base de données.

- 4 Dans tous les cas, l'enjeu est clairement de travailler, dans le cadre d'un projet qui relève du prototype, sur la possibilité de corpus ouvert, les bases et les documents textuels reliés ayant vocation à être augmentés de sources qui ont des structures, des langues et des natures différentes. C'est la raison pour laquelle il n'est pas pour nous question ici de présenter une seule et unique mise en œuvre numérique possible, mais de partager une solution parmi toutes celles qui sont réalisables, et d'en évaluer les atouts mais aussi les limites techniques. Il convient en effet de garder à l'esprit que l'élaboration des plans de projets numériques – au-delà des restrictions scientifiques – reste inévitablement limitée aux compétences techniques tant des chercheurs que des ingénieurs impliqués, dans un contexte qui impose par essence la transdisciplinarité<sup>5</sup>.

## **Insertion de la dimension numérique dans un projet comme *PoBLAM* : quelques points de vigilance**

### **Logiciels *open source* et interopérabilité des données**

- 5 Faisant partie des disciplines mouvantes, les humanités numériques doivent se confronter aux innovations numériques constantes et obligent les chercheurs à allier leur compétence disciplinaire à une conscience réelle de l'évolutivité rapide des solutions technologiques, mais sans jamais perdre de vue que les projets doivent atteindre, pour perdurer, une forme de stabilité. Mais comment un projet comme *PoBLAM*, dépendant en partie de technologies numériques, mouvantes et constamment évolutives, peut-il prétendre à une forme de stabilisation ? Autrement dit, se pose la question bien connue : quels sont les langages et outils qui permettent d'obtenir le résultat le plus stable possible et, à l'inverse, à quel moment faut-il cesser d'innover technologiquement dans un projet pour pérenniser l'existant et sous quelle forme ?
- 6 Chaque compétence technique connaît un processus de développement dont la durée de vie est intimement liée à son inéluctable obsolescence. Certains facteurs de choix accompagnent plus aisément que d'autres l'évolution de la donnée. L'utilisation de formats libres sans restriction d'accès et de langages publics en fait à nos yeux partie car ils sont souvent très bien documentés et suffisamment employés pour assurer leur maintenance ou leur migration en cas d'obsolescence<sup>6</sup>. En effet, situer notre projet dans la science ouverte passe non seulement par l'utilisation de langages et d'outils *open source* et interopérables mais impose aussi de le construire en visant dès le début une valorisation pérenne des données sous des formats aisément partageables et pouvant être ensuite adaptés à l'évolution d'outils de visualisation sans toucher à l'intégrité même des données créées. En ce sens, les outils de visualisation sont conçus à la fois comme du consommable (susceptibles d'être repensés dans d'autres contextes) et du pérenne dans la mesure où les concepteurs du projet peuvent imposer (sur des critères de rationalité scientifique) telle procé-

de visualisation en excluant (car non scientifiquement pertinente ou introduisant un trop grand biais) d'autres visualisations pourtant techniquement possibles.

- 7 Il s'agit donc, dès l'ouverture du projet, de trouver un juste équilibre entre les problématiques liées à l'innovation permanente des techniques et la stabilité dans la transmission pérenne des données. Cette transmission pérenne des données implique de pouvoir les utiliser pour construire des connaissances, les structurer et les encoder de façon rationnelle, et, enfin, les visualiser en s'interrogeant sur les publics qui seront amenés à consulter ces données.

## Réflexion scientifique et pratique numérique : un ensemble de compromis

- 8 L'utilisation du numérique a bouleversé notre manière de concevoir la recherche dans le sens où les données, bien que pensées et structurées en lien avec les modalités scientifiques des cultures académiques, sont contraintes, en même temps d'ailleurs qu'éclairées, par la technique que ce soit par les limites de l'outil que par les capacités techniques des chercheurs ; en effet, dans le cas de la construction de nos données spécifiques qui reposent sur le croisement de paramètres nombreux, la difficulté technique peut certes contraindre leur forme (« on ne sait pas faire » ou, plus souvent, « je ne sais pas faire ») mais elle peut aussi questionner la légitimité de telle structure de données (« ai-je vraiment scientifiquement besoin de repousser ici les limites de ce que propose tel ou tel outil ? »). Finalement, c'est tout un changement dans le procédé d'agencement de la recherche qui est en œuvre, puisqu'aux difficultés scientifiques traditionnelles s'ajoutent celles de la technique et le piège du « déterminisme technologique » (Sinatra et Vitali-Rosati 2014). Sur ce point, il serait naïf et trompeur de prétendre que cela n'a jamais influencé les projets numériques de *PoBLAM*, puisque cela reste une conséquence directe et involontaire liée à la difficile transdisciplinarité dans ces domaines de recherche. Il est indispensable de mêler, dès la conception, la réflexion scientifique disciplinaire, la pratique technique et les notions réflexives qui entourent les humanités numériques – soit quoi faire, comment faire et pourquoi – pour pondérer les trois domaines de manière rationnelle en accord avec les traditions savantes<sup>7</sup>, et ainsi ne pas se laisser conduire sur un chemin dangereux où les technologies supplantent l'application humaine jusqu'à une tendance à la performance plus qu'à la pertinence ou l'intérêt.

- 9 En effet, un philologue (dans le cas de nos projets en l'occurrence) qui n'a aucune idée des technologies des humanités numériques aura du mal à formuler son exacte demande technique, voire ne verra même pas qu'il peut formuler une demande technique, tandis que l'ingénieur, s'il n'a aucune vue de la discipline sur laquelle il va devoir collaborer avec le chercheur, pourra ne pas comprendre l'attente exacte de ce chercheur ou imposer une limite à la recherche disciplinaire en fonction de ses compétences ou goûts techniques. Ainsi, tout projet impliquant des humanités numériques comporte obligatoirement un volet appliqué et les extrapolations ou généralisations que l'on peut tirer de notre travail doivent être nuancées des nécessaires conditionnements imposés à notre usage de la technique par la nature même de la discipline concernée (ici la philologie classique, l'histoire des textes et des idées).

## L'éditorialisation au service de la transmission des données

- <sup>10</sup> Visualiser des données, c'est penser aux principes d'éditorialisation (Sinatra et Vitali-Rosati 2014<sup>8</sup>), autrement dit, c'est réfléchir aux moyens technologiques mis au service de ce que l'on veut présenter, en gardant à l'esprit que, en retour, ces outils déterminent en partie les possibilités de réalisation.
- <sup>11</sup> La transmission des données sur un projet aussi varié dans ses aspects que *PoBLAM* doit être réfléchi à différents moments de l'élaboration numérique pour que la visualisation se plie à la nature du corpus et non l'inverse. Pour la conception d'un hypertexte ou d'une édition numérique savante, l'enjeu sera de déterminer les bonnes balises à appliquer en fonction des éléments que l'on souhaite distinguer sur l'ensemble du corpus, en prenant d'ores et déjà en compte la structure que devront avoir les données, en vue de leur exploitation dans une base. Mais il n'est pas question pour autant d'oublier que, dans une deuxième phase, ces données seront interprétées en vue d'une visualisation spécifique, puisqu'aux choix liés à la nature scientifique du projet s'ajouteront possiblement d'autres informations numériques qui aideront à leur exploitation ultérieure. Le plus important est ici que les choix de structuration des données et de balisage mis en place spécifiquement pour la visualisation n'altèrent pas les choix scientifiques. Ces choix de visualisation viennent en complément en tant qu'interface de lecture entre chercheur et interlocuteur pour répondre aux besoins variés des différents publics visés sans modifier les objectifs scientifiques des projets.
- <sup>12</sup> Toutefois, lorsque l'on parle de visualisation, on ne pense strictement qu'à celle dont l'usage se limite à la consultation des données, alors qu'au-delà de la visualisation évidente consacrée à leur diffusion, il faut savoir réfléchir à la mise en place d'une interface technique servant d'intermédiaire entre l'ingénierie de recherche et les chercheurs. Par exemple, dans le cas de notre équipe dont certains membres sont novices ou étrangers au domaine des humanités numériques, on doit penser à une manipulation simplifiée de l'enrichissement du corpus, car les intervenants scientifiques risquent de perdre du temps et de se décourager si on leur propose un mode de saisie trop complexe. Ainsi, l'utilisation de commentaires détaillés sur les fichiers d'encodage textuel et la création de raccourcis dans les logiciels de base de données relationnelle ont été, comme nous le verrons plus tard, d'une grande aide pour conditionner un enrichissement qualitatif des corpus. La rédaction de manuels spécifiques nous aide à préserver à la fois une maintenance cohérente mais aussi une transmission sereine des données pour la suite de *PoBLAM* puisque les projets seront, nous l'espérons, amenés à évoluer. Une documentation précise est indispensable à leur continuité tout en participant au partage des techniques numériques, dont nous allons maintenant explorer la diversité.

## Perspectives d'encodage structurel : du texte à la donnée et vice versa

- <sup>13</sup> Comme nous l'avons vu, les projets numériques de *PoBLAM* relèvent essentiellement de deux types, selon qu'ils sont orientés textes ou orientés données.

- 14 Dans le premier cas, il s'agit de réaliser des éditions numériques de textes, peu ou mal édités, accompagnés le cas échéant de traductions, et surtout augmentés de diverses possibilités d'interrogation, par personnage cité, par lieu, par auteur utilisé (intertextualité avouée ou implicite), par liste des références à d'autres textes, etc. De plus, s'agissant parfois de littérature secondaire (autrement dit de commentaires d'un autre texte), il devient nécessaire de fournir le texte commenté en même temps que le texte du commentaire, et de permettre la navigation parallèle entre les deux entités.
- 15 Pour les projets orientés base de données, il est évident que, si certains sont purement pensés et construits dans une optique de structuration de données, d'autres sont à la marge avec des objectifs d'édition numérique : par exemple, lorsque dans les bases de données intertextuelles, on veut pouvoir accéder à la lecture plein texte du texte source à partir du repérage d'un passage cité (notamment, pouvoir lire tout un poème dont un poète ne citerait qu'un vers), l'outil de gestion de base de données doit dialoguer avec un outil d'édition numérique en tenant compte des spécificités de chacun ; cela influe à la fois sur la structure de la base de données (qui doit être pensée dans une orientation texte) et sur l'encodage du texte lui-même, en fonction de la manière dont il va se trouver lié à la base de données.
- 16 Ainsi, il nous a semblé évident d'adopter une approche d'encodage des textes en XML-TEI, du fait de sa polyvalence et de sa prédominance dans les domaines de la recherche en sciences humaines et sociales et des diverses possibilités offertes par le langage de transformation XSLT. Cette technique reste la plus appropriée pour nos travaux car c'est un langage nativement textuel et c'est donc sans surprise que nous l'avons utilisé pour *Poéxégèse* et *Barbosa*, avec la possible valeur ajoutée de faire dialoguer ces projets et d'autres du même type qui pourraient émerger dans notre réseau. Cela nous a posé davantage de difficultés pour la gestion de *Vestigia Iuuenci*, qui est orienté à la fois textes et données et dans lequel il nous fallait une interface de fouille avec un système de filtre qu'il est impossible de réaliser en XSLT. C'est pourquoi, nous avons choisi d'élaborer une base de données MySQL, alimentée grâce aux fichiers textuels XML, qui offre une alternative aisée face aux contraintes techniques, tout en favorisant une nouvelle interprétation numérique de l'ensemble.

## L'encodage textuel XML-TEI

- 17 Les deux projets qui ont posé de la manière la plus décisive la question du choix des modalités d'encodage en XML-TEI sont *Barbosa* et *Poéxégèse*.

### L'édition du commentaire d'Ayres Barbosa à l'*Historia Apostolica* d'Arator

- 18 Dans le cadre de l'axe d'étude sur la diffusion des poètes chrétiens dans le milieu scolaire, que ce soit par les recueils de textes ou les commentaires, une partie du travail consiste en l'édition numérique « augmentée » du commentaire d'Ayres Barbosa à l'*Historia Apostolica* d'Arator, publication en 1516 d'une série de cours que le grand érudit portugais avait donné à l'université de Salamanque<sup>9</sup>. Disponible en version numérisée, l'ouvrage de Barbosa se présente dans une mise en pages assez fréquente pour les textes de cette époque : le texte du poème se trouve dans la zone centrale de la double page et il est entouré par un commentaire extrêmement volumineux puisqu'il peut aller jusqu'à plus de dix fois le volume du texte commenté (figure 1).

Figure 1. Un exemple de page du commentaire d'Ayres Barbosa

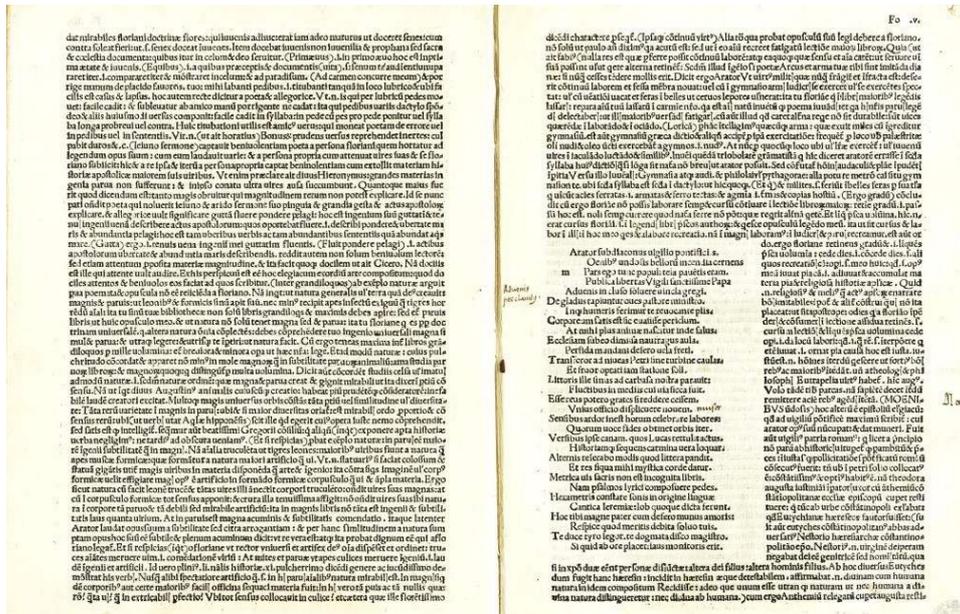


Image produite par les auteurs · Capture d'écran effectuée sur logiciel Foxit-Reader

19 De fait, le projet n'a pas prioritairement pour but de fournir une analyse du texte de Barbosa, et encore moins un commentaire, mais de donner le matériau nécessaire à un travail sur le texte et à des recherches ultérieures ; à ce titre, il s'agit bien d'édition au sens strict, mais dans la logique d'une édition augmentée.

20 La première étape a été d'obtenir, avec un outil HTR (*handwritten text recognition*), un fichier XML-TEI de l'ensemble du volume, en séparant ensuite en deux fichiers le texte commenté et le commentaire ; il aurait évidemment été possible de conserver la mise en pages originale et donc de publier ensemble texte et commentaire dans une disposition approuvante du fac-similé, mais ce choix n'aurait rien apporté de décisif, s'agissant d'un texte imprimé et qui, abréviations usuelles au début du XVI<sup>e</sup> siècle mises à part, ne présente aucune difficulté de lecture. L'outil eScriptorium<sup>10</sup>, choisi pour cette reconnaissance du texte, a donné, après entraînement, de très bons résultats, ne nécessitant d'intervention lourde sur le résultat que dans deux cas : lorsque la page est abîmée et donc le texte difficilement interprétable, même après entraînement, et lorsque le système abrégatif change en cours de route. Le commentaire étant extrêmement long, plusieurs typographes se sont relayés pour le mettre en page, sans harmoniser leur pratique des abréviations, ce qui produit une série de confusions difficilement évitables.

21 Ce texte pivot a été ensuite structuré en <div> en respectant la structure originale (*prima, secunda, tertia*) indiquant chaque division du poème et la partie de commentaire correspondante, mais en ajoutant une division plus fine : le texte du commentaire n'étant pas pourvu de paragraphes, et la ponctuation (point ou deux-points et, plus rarement, point d'interrogation) ne permettant pas d'isoler les phrases, il a fallu restructurer le texte latin en utilisant la balise <seg> de manière à pouvoir isoler des unités de sens qui pourraient ensuite être appelées lors de l'enrichissement de la lecture du texte (voir ci-dessous). Toutefois, la ponctuation, ainsi que la graphie, faisant partie de l'édition originale, a été encodée avec attribution à Barbosa, sachant bien qu'elle est sans doute plus le résultat du travail de ses imprimeurs que de son écriture ; elle se présente donc sous la forme <pc resp="Barbosa">, ce qui permet de reponctuer si besoin le texte selon les normes de la ponctuation moderne en variant simplement le @resp.

22 Pour le texte du poème, traité selon les mêmes divisions (au <div> correspondent des <lg>), tout l'enjeu était de pouvoir relier texte et commentaire selon une procédure permettant d'appeler le vers commenté, pour illustrer le commentaire, et le commentaire, pour illustrer le vers commenté. Les citations du poème dans le commentaire ont été encodées sous la forme de <quote type="lemma"> avec un @corresp renvoyant à la numérotation des vers selon une forme normalisée (figure 2).

Figure 2. Représentation d'un lemme dans le commentaire (a), puis dans le fichier poème (b) et dans la sortie HTML (c)

a) <quote type="lemma" corresp="HA\_C1\_56">sedebat maria</quote><pc resp="Barbosa">:</pc>  
 b) <lg>Porta Maria dei<pc resp="Barbosa">.</pc> genitrix intacta creantis<pc resp="Barbosa">.</pc></lg>  
 c) <span class="line" id="C1\_57"><span class="line\_number">57: </span>Porta Maria dei.  
 genitrix intacta creantis.<br></span>

Image réalisée par les auteurs · Capture d'écran effectuée sur le logiciel Oxygen

23 Il était important de conserver ce lien avec le poème tel que Barbosa le lit car le texte qu'il transmet n'est ni exactement celui des éditions modernes, ni même celui des éditions contemporaines. De plus, le commentateur ne commente parfois pas le texte qu'il édite, mais une variante dont on ignore l'origine, ce qui impose l'utilisation de balises du type <app> pour indiquer ces divergences.

24 Pour ce qui est de l'annotation du texte même du commentaire, plusieurs éléments ont retenu l'attention : outre les habituelles entités nommées (toponymes, anthroponymes) indispensables à la construction d'indices et à l'exploration du texte, le commentaire présente des singularités intéressantes dans la perspective de leur traitement numérique : de très nombreuses références à des textes bibliques, patristiques, scolastiques ou humanistes émaillent le propos, mais ces citations sont assez souvent anonymes (« comme l'a dit quelqu'un ») ou imprécises (« Jérôme dans sa lettre <sup>11</sup> »), voire fausses car reposant sur un empilement de citations qui transmet l'erreur. Par exemple, Barbosa peut attribuer à Augustin un texte soit d'un autre Père, soit de source inconnue, parce qu'il a trouvé cette attribution dans un passage du scolastique Alexandre de Halès. Dans ce cas, un triple niveau d'annotation est nécessaire pour rendre compte de la complexité des données en indiquant que le texte n'est pas d'Augustin, qu'il l'est néanmoins aux yeux de Barbosa, et que l'erreur vient de sa source. L'encodage prend alors les formes suivantes : pour le type simple, <persName nymRef="#Augustinus" ana="Bonaventura">Augustinus</persName> (sachant que « Augustin » est en fait une autre personne) ; pour le type complexe, la solution de la figure 3.

Figure 3. Représentation d'une référence complexe (erronée)

```
<bibl type="scriptores_medii_aevi" ana="Petri_Lombardi_Sententiae">
  <persName nymRef="#Augustinus" ana="Petrus_Lombardus">
    augustinus
  </persName>
</bibl>
```

Image réalisée par les auteurs · Capture d'écran effectuée sur le logiciel Oxygen

25 De plus, il importe, comme dans tout texte de commentaire, de signaler ce qui relève de l'emploi en mention (pour définir tel ou tel mot, par exemple) et de sourcer les références en utilisant un système référentiel connu de la communauté des lecteurs (ici le *Thesaurus Linguae Latinae* [TLL] quand il existe) sur le modèle simple suivant : <cit

source="#Aen\_1\_39"><sup>12</sup>. Tout ce travail d'encodage aboutit ensuite à l'extraction automatique avec XSLT de ces données vers un fichier listes (listPerson, listPlace, listNym, listBibl) où les références sont normalisées et redirigées vers des pages extérieures permettant de les définir ou d'y accéder directement, sur le modèle, par exemple, de <person role="persona\_historica" xml:id="Agostino\_Barbarigo" source="https://www.treccani.it/enciclopedia/agostino-barbarigo">. Le choix ici est de faciliter l'identification des entités et, dans le cas de textes, de pouvoir renvoyer à la lecture du texte intégral s'il est disponible en ligne, ce qui d'ailleurs n'est pas toujours le cas. Cette extraction vient s'insérer dans la constitution de jeux de données, en lien avec les autres projets, comme *Vestigia Iuueni*, dans une optique de fusion à terme de tous les objets numériques dans un portail commun. Il nous a semblé très important de générer ces données indépendamment de leur visualisation (alors qu'on aurait pu tout aussi bien les visualiser directement depuis le fichier pivot XML-TEI, précisément pour articuler ce projet avec des projets plus orientés sur les données) ; en un sens, nous faisons ici le chemin du texte vers la donnée, qui est inverse de celui de *Vestigia Iuueni*, comme nous le verrons.

26 Tous ces éléments d'encodage n'affectent que le texte latin, et non la traduction réalisée sur un fichier parallèle relié au texte du commentaire par la numérotation identique des <seg>, ce qui permet ensuite la parallélisation (figure 4).

Figure 4. Encodage parallèle du texte (a) et de la traduction (b) au niveau du segment

a) <seg n="707">  
 (<quote type="lemma" corresp="HA\_C1\_56">Qua tunc  
 statione</quote><pc resp="Barbosa">.</pc> statio locus dicitur in quo  
 b) <seg n="707"> (Qua tunc statione). On appelle « statio » un endroit dans lequel  
 navires ou soldats séjournent pour un temps, ou bien un lieu assigné à la garde,

Image réalisée par les auteurs · Capture d'écran effectuée à partir du logiciel Oxygen

27 Le choix de ne pas encoder la traduction repose sur le constat qu'une traduction (ici en français) a pour but d'être mise éventuellement en regard d'autres traductions et donc de ne pas exister autrement qu'en appui à l'encodage scientifique qui ne repose donc que sur le latin.

28 La visualisation des éléments s'opère d'abord directement, avec un mode de lecture en ligne du commentaire en langue originale, un mode de lecture du poème et un mode de lecture du commentaire accompagné de sa traduction en regard. Dans tous ces systèmes, toutes les entités nommées sont rendues cliquables et permettent d'accéder à une page qui présente toutes les occurrences de cette entité en contexte (au niveau <seg>) avec possibilité de retourner dans le commentaire en partant d'une autre occurrence par un lien hypertextuel.

29 Un autre mode est celui de la visualisation par listes, classées par type et ordre alphabétique des éléments index personarum, index locorum, index librorum, etc., ouvrant sur le même type d'affichage (figures 5, 6, 7 et 8).

Figure 5. Barre de menus du site (version provisoire)



Image réalisée par les auteurs

Figure 6. Barre de menus d'accès au poème



Image réalisée par les auteurs

Figure 7. Visualisation du texte et de sa traduction

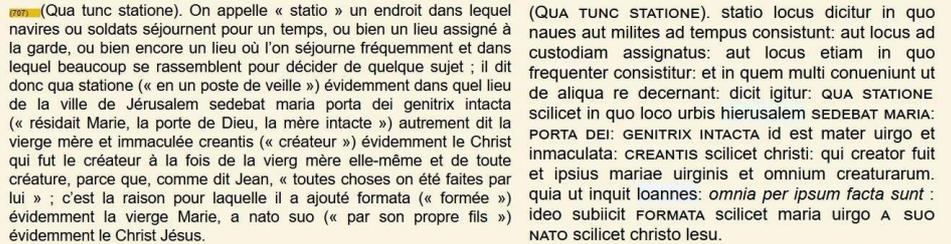


Image réalisée par les auteurs

Figure 8. Exemple de recherche sur un nom de personnage

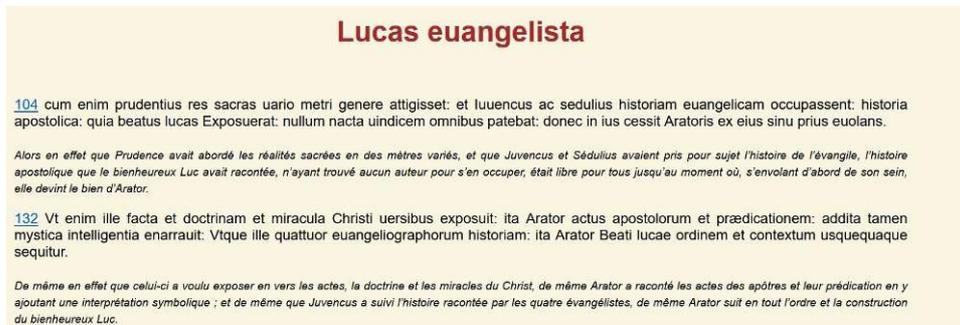


Image réalisée par les auteurs

30

Un tel projet doit évidemment beaucoup à d'autres projets, comme *Hyperdonat*<sup>13</sup>, développé par C. Nicolas et B. Bureau, et dont le principe est substantiellement le même, mais en tenant compte de nouvelles exigences, par exemple la mise en rapport des entités avec des éléments extérieurs (cartes, encyclopédies, corpus, etc.), pour répondre aux attentes d'une analyse du texte.

Figure 9. Diagramme de production du projet Barbosa

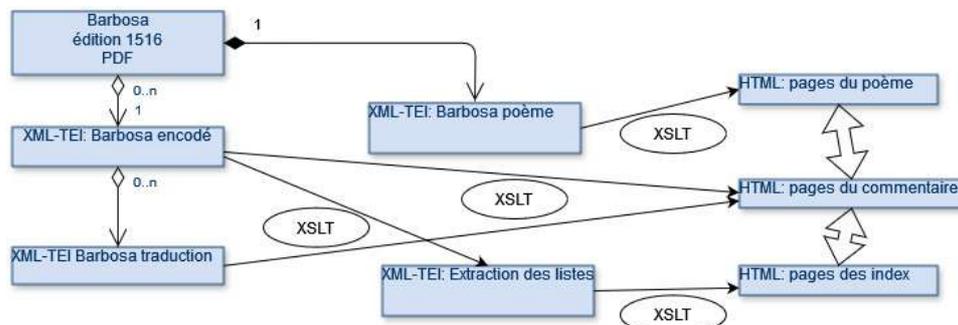


Image réalisée par les auteurs, créée à partir du logiciel Draw Io

## Le projet hypertexte *Poéxégèse*

31

*Poéxégèse*<sup>14</sup> est un projet porté par Francesco Stella ayant pour vocation d'identifier dans des poèmes médiévaux la trace explicite ou implicite de citations bibliques, non seulement dans les cas directs (tel poète cite tel passage de la Bible), mais encore dans les cas indirects (tel poète cite tel

commentateur qui cite tel passage biblique), ce qui complique considérablement le traitement des éléments. L'objectif était donc de créer un hypertexte à partir de la lecture des textes bibliques avec un regroupement des diverses références – directes et indirectes – issues des commentaires et des textes poétiques. Le projet, qui était dans sa phase initiale, s'est concentré sur un seul groupe limité de livres bibliques, les *Libri Regum* et leurs commentaires, répertoriés au format XML dans le *Corpus Corporum*<sup>15</sup>, mis en rapport avec trois poètes médiévaux, Hildebert de Lavardin, Bernard de Cluny et celui que l'on nomme Pierre de Riga, sans préjuger de l'auctorialité exacte du texte (Dinkova-Bruun 2019).

32 La première étape dans la collecte du corpus a été d'harmoniser l'encodage de la structure des textes et de leur numérotation en fonction de leur nature afin qu'ils puissent communiquer ensemble numériquement. Nous avons donc appliqué un identifiant unique et explicite pour chaque élément qu'il serait utile de cibler par la suite : les chapitres et versets pour les textes de la Bible, les livres et les vers pour la poésie, et les livres et paragraphes pour les commentaires. Ces identifiants ont été utilisés comme liens hypertextes de références et de citations depuis les textes citateurs – les textes poétiques et commentaires – vers les passages bibliques cités. Cette étape-clé a été la plus chronophage puisque nous n'avons mis en place aucune automatisation d'encodage ; ainsi, chaque référence a été ajoutée et validée à la main (figure 10).

Figure 10. Appel de référence de 1S, 1, 2 (abrégé S1\_1\_2) chez Raban Maur

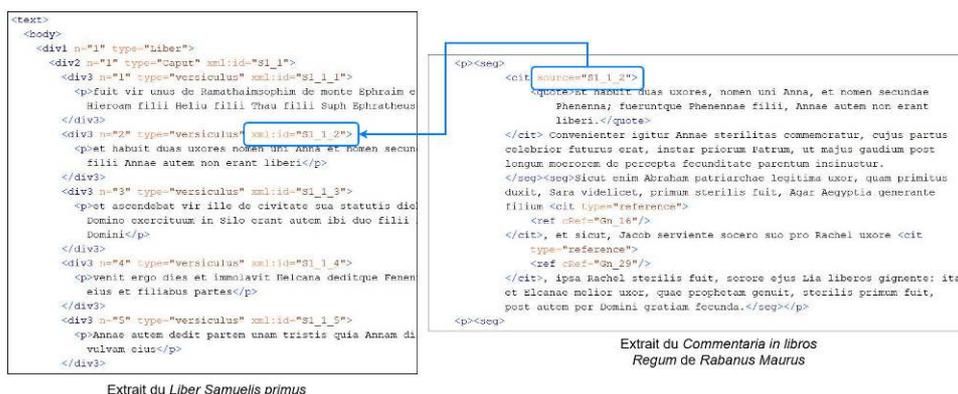


Image produite par les auteurs · Capture d'écran effectuée sur le logiciel Oxygen

33 Pour visualiser ces multiples connexions et obtenir un hypertexte, nous avons transformé les fichiers XML en fichiers HTML au moyen de feuilles de transformation XSLT. L'élément central est un outil de comparaison partant des textes de la Bible, depuis une navigation au verset, suivi d'un affichage des références et citations recensées dans les textes poétiques et les commentaires.

Figure 11. Résultat de l'outil de comparaison 1 S, 1, 2 où sont encadrés en bleu les éléments d'encodage précédemment illustrés

**LIBER PRIMUS SAMUELIS**

**Caput 1**

Liens vers les chapitres : [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [23](#) [24](#) [25](#) [26](#) [27](#) [28](#) [29](#) [30](#) [31](#)

Liens vers les versets : [1](#) [2](#) [3](#) [4](#) [5](#) [6](#) [7](#) [8](#) [9](#) [10](#) [11](#) [12](#) [13](#) [14](#) [15](#) [16](#) [17](#) [18](#) [19](#) [20](#) [21](#) [22](#) [23](#) [24](#) [25](#) [26](#) [27](#) [28](#)

Précédent V. 2 : et habuit duas uxores nomen uni Anna et nomen secundae Fenenna fueruntque Fenennae filii Annae autem non erant liberi Suivant

Poètes	Commentaires
<p>HILDEBERTUS CENOMANENSIS 1056-1133</p> <p><i>liber 1, v. 1. 2. :</i></p> <p>Helcana de Ramatha Phenennae sponsus et Annae</p> <p>Venerat in Silo sacra Deo facere.</p> <p>PETRUS RIGA</p> <p><i>liber 1, v. 23. 24. :</i></p> <p>Anna, Phenenna, due soli nupsere marito;</p> <p>Hec sterilis partu, fertilis illa fuit.</p> <p><i>liber 1, v. 23. 26. 27. 28. :</i></p> <p>Gratia dicitur hec, conuersio dicitur illa;</p> <p>Suscipit ista fidem, deserit illa malum;</p> <p>Primo Phenenna parit quia te conuersio mutans</p> <p>Dat primos fructus ut mala prima gemas.</p> <p><i>liber 1, v. 49. 50. 51. 52. :</i></p> <p>Que fuerat sterilis prius et confusa dolore</p> <p>Fit fecunda, Deo cantica leta canit:</p> <p>Sic agit ecclesia; sterilis fuit ante frequentans</p> <p>Idola, nunc fructus parturit illa bonos.</p> <p>BERNARDUS CLUNIACENSIS</p> <p><i>liber, v. 3. :</i></p> <p>Anna sibi sterilis uxor, fecunda Fenenna.</p>	<p>GREGORIUS I (540-604)</p> <ul style="list-style-type: none"> <li>- Proemium 1, Caput 1</li> <li>- Liber 1, Caput 1</li> <li>- Liber 2, Caput 1</li> </ul> <p>ANGELOMUS LUXOVENSIS(-c.895)</p> <ul style="list-style-type: none"> <li>- Liber 1, Caput 1</li> </ul> <div style="border: 2px solid blue; padding: 5px; margin: 5px 0;"> <p>RABANUS MAURUS (780-856)</p> <ul style="list-style-type: none"> <li>- Liber 1, Caput 1</li> </ul> <p>« <i>Et habuit duas uxores, nomen uni Anna, et nomen secundae Phenenna; fueruntque Phenennae filii, Annae autem non erant liberi.</i> » ( S1,1,2) Convenienter igitur Annae sterilitas commemoratur, cuius partus celebrior futurus erat, instar priorum Patrum, ut majus gaudium post longum moerorem de percepta fecunditate parentum insinnetur. Sicut enim Abraham patriarchae legitima uxor, quam primitus duxit, Sara videlicet, primum sterilis fuit, Agar Aegyptia generante filium (voir: Ga, 16), et sicut, Jacob serviente socio suo pro Rachel uxore (voir: Ga, 29), ipsa Rachel sterilis fuit, sorore ejus Lia liberos gignente: ita et Elcanae melior uxor, quae prophetam genuit, sterilis primum fuit, post autem per Domini gratiam fecunda.</p> </div> <p>THEUTMIRUS; CLAUDIUS TAURINENSIS</p> <ul style="list-style-type: none"> <li>- Proemium, Caput 3</li> <li>- Liber 1, Caput 1</li> </ul>

Image produite par les auteurs : [http://poexegese-girpam.u-strasbg.fr/S1\\_1\\_2.html](http://poexegese-girpam.u-strasbg.fr/S1_1_2.html)

34 Il est également possible de lire les textes tels quels, et de se diriger vers l'outil de comparaison uniquement pour les éléments qui attirent notre curiosité à un moment précis de la lecture. Les liens hypertextes sont accessibles au niveau des numéros des versets pour les textes de la Bible, en bas de chaque unité phrasée pour les poésies et à l'endroit précis de la référence ou de la citation dans les commentaires.

35 L'analyse textométrique du corpus faisait également partie des prérequis du projet ; nous avons ainsi intégré à chaque page de lecture des textes un lien permettant le téléchargement des fichiers XML, laissant la possibilité aux utilisateurs plus aguerris de lemmatiser le corpus et d'utiliser librement des logiciels d'analyse puissants tels que TXM ou BaseX. Nous voulions malgré tout ouvrir ce champ d'étude aux plus novices en la matière et proposer une fouille des données automatisée sur des thématiques prédéfinies à partir des textes intégrés à *Poéxégèse* : le corpus complet, un certain nombre de textes de la Bible, les textes poétiques et enfin les commentaires. À cet effet, nous avons transféré les fichiers XML directement dans Voyant Tools<sup>16</sup> et intégré les liens redirigeant vers l'outil dans le site Web du projet<sup>17</sup>. Cette première approche permet aux personnes moins familières de ces techniques de s'exercer tout en laissant la possibilité de créer un corpus d'analyse propre en important des fichiers XML depuis le site *Poéxégèse* vers l'outil Voyant Tools (figure 12).

Figure 12. Analyse du mot *dominus* dans le corpus complet sur Voyant Tools

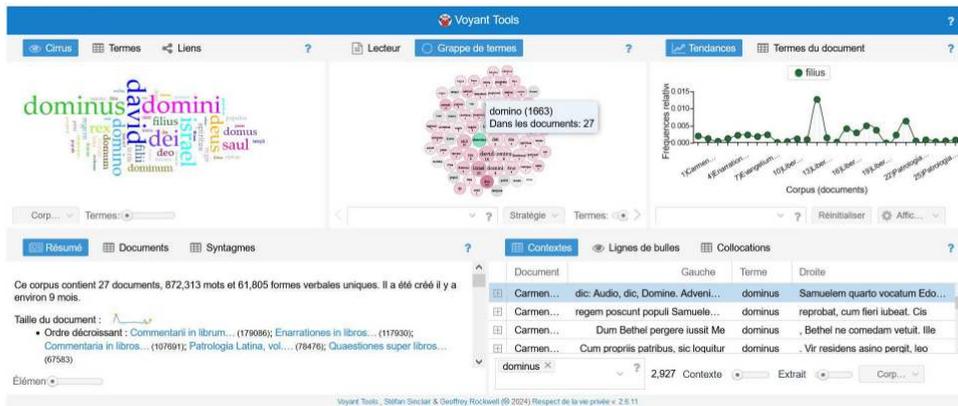


Image produite par les auteurs · Capture d'écran effectuée sur le site Voyant Tools

- 36 L'avantage principal de cette méthode réside dans le simple enrichissement futur de la base de données, puisqu'il suffit de suivre la cohérence (1) du choix des identifiants numériques des passages de la Bible et (2) des choix de balises dans l'encodage des références et des citations dans les textes de la Bible et des commentaires.

## Un projet mixte d'édition et de base de données : *Vestigia Iuueni*

- 37 Le projet *Vestigia Iuueni* nous a été présenté par Stefan Freund et Philip Buckl de l'université de Wuppertal dans le but de construire l'histoire des citations ou allusions au poète Juvencus (vers 325) dans l'œuvre de plusieurs poètes ultérieurs<sup>18</sup>. Cette base de données n'a pas pour vocation première de transmettre une analyse scientifique complète pour chaque intertextualité, ni d'automatiser leur référencement, mais de mettre à disposition leur compilation sous la forme la plus exhaustive possible, en incluant donc un grand nombre de métadonnées. Le plus important était donc de développer une interface simple et efficace pour une consultation aisée des données et ainsi faciliter leurs analyses.
- 38 Nous avons deux contraintes techniques impératives qui étaient (1) de pouvoir accéder aux informations en naviguant à travers la lecture des textes – de Juvencus vers les auteurs et des auteurs vers Juvencus – et (2) de fouiller l'ensemble des données avec un outil de filtre de recherche le plus simple possible.
- 39 Il aurait sans doute été possible de réaliser ce projet au moyen d'un encodage complet du corpus en XML-TEI, comme pour *Poëxégèse* et *Barbosa*, et de créer l'interface de visualisation à l'aide de JavaScript. Toutefois, pour éviter plusieurs mois de tentatives de programmations complexes qui, par ailleurs, auraient pu ne jamais aboutir, nous avons privilégié le choix d'un SGBD (Système de gestion de base de données) déjà construit et élaboré précisément pour donner aux chercheurs, sans trop de difficultés d'apprentissage, le contrôle sur l'intégration et la gestion de leurs données.
- 40 Nous avons opté pour la plateforme *open source* Heurist<sup>19</sup> car elle dispose de multiples avantages. Elle est hébergée et soutenue par Huma-Num qui garantit son maintien et finance son développement. Son créateur, Ian Johnson, et une communauté active<sup>20</sup> assurent un soutien et dispensent des conseils techniques aux utilisateurs<sup>21</sup>. Mais c'est surtout la possibilité d'importer les données textuelles en CSV et d'exporter les informations des intertextualités renseignées dans Heurist au format XML-HML<sup>22</sup> qui

nous a définitivement confortés dans notre choix. Grâce à cela, nous avons mis en place un plan de projet articulé en deux composantes essentielles : la réalisation de la base de données et la création d'un hypertexte, où l'on pourrait naviguer entre les mots du corpus qui font l'enjeu d'une ou de plusieurs intertextualités.

41 L'entité minimale élémentaire pour notre étude étant le mot – c'est cette donnée qui rend possible la publication des textes, le renseignement des intertextualités dans Heurist et qui permet la liaison entre les deux –, la première étape a été de tokéniser nos fichiers XML en distinguant tous les mots et en leur donnant à chacun un identifiant évocateur.

Figure 13. Extrait du fichier XML et HTML sur les trois premiers vers de l'*Hamartigeneia* de Prudence

Figure 13 displays two side-by-side screenshots. The left screenshot shows XML code for the first three verses of the *Hamartigeneia* by Prudentius. The XML uses a hierarchical structure with `<div1 type="Praefatio">` and `<h1>` tags. Individual words are enclosed in `<w n="...">` tags, where 'n' is a unique identifier. The right screenshot shows the corresponding HTML rendering. It features a main title **Hamartigeneia, Prudentius** and a sub-section **Praefatio**. Below this, three entries are listed, each starting with a bolded identifier (e.g., **00001**) followed by a list of words in parentheses, such as `(#PruHam00001) ephebi (#PruHam00002) fossor (#PruHam00003) et (#PruHam00004) pastor (#PruHam00005) duo (#PruHam00006)`.

Image produite par les auteurs · Capture d'écran du code XML effectuée sur le logiciel Oxygen

42 Tel que vous pouvez le voir dans l'illustration ci-dessus (figure 13), nous avons adopté une normalisation des identifiants avec les trois premières lettres de l'auteur<sup>23</sup>, les trois premières lettres de l'œuvre<sup>24</sup> et, enfin, un numéro lié à sa localisation exacte sous un format en cinq chiffres<sup>25</sup>. De cette manière, nous empêchons d'établir une quelconque confusion entre les mots similaires du corpus tout en conservant un lien invariable entre les fichiers XML-TEI et la base de données Heurist. Pour garantir l'enrichissement de la base de données, nous avons fourni aux chercheurs les fichiers HTML de tous les textes qui renseignent l'identifiant de chaque mot. On évite les erreurs puisqu'ils n'ont plus qu'à le copier dans la barre de recherche et de sélection du terme concerné. En parallèle, nous avons encodé la structure de nos textes en numérotant les livres et les vers, ce qui garantit à la fois une cohérence globale et une connexion fluide entre les mots et d'autres éléments de l'œuvre.

43 Une transformation XSLT de XML-TEI vers CSV nous a permis d'importer toutes les données dans des tables dédiées aux auteurs, aux textes et à l'intégralité de leur contenu et de basculer nos textes en une suite de données discrètes et structurées<sup>26</sup> (figure 14). Car l'un des grands atouts d'Heurist réside dans sa capacité à importer automatiquement et facilement des quantités très importantes de données. Cette fonctionnalité s'est avérée indispensable lorsque nous avons dû modifier certains identifiants en cours de projet et pour lesquels nous avons simplement importé un nouveau fichier CSV contenant les anciennes et nouvelles informations, ce

qui a permis d'actualiser l'existant sans altérer l'intégrité de la base. Et cela même sur plus de 100 000 entrées différentes, ce qu'il aurait été impossible de faire manuellement.

Figure 14. Extrait du fichier GSV pour l'importation des données textuelles des trois premiers vers de l'*Hamartigeneia* de Prudence

	A	B	C	D	E	F	G
1	Author	Work	Books	Number	Text line	Number of the word	Text of the word
18957	Prudentius	Hamartigeneia	Praefatio	00001	Fratres ephebi fossor et pastor duo	PruHam00001	Fratres
18958	Prudentius	Hamartigeneia	Praefatio	00001	Fratres ephebi fossor et pastor duo	PruHam00002	ephebi
18959	Prudentius	Hamartigeneia	Praefatio	00001	Fratres ephebi fossor et pastor duo	PruHam00003	fossor
18960	Prudentius	Hamartigeneia	Praefatio	00001	Fratres ephebi fossor et pastor duo	PruHam00004	et
18961	Prudentius	Hamartigeneia	Praefatio	00001	Fratres ephebi fossor et pastor duo	PruHam00005	pastor
18962	Prudentius	Hamartigeneia	Praefatio	00001	Fratres ephebi fossor et pastor duo	PruHam00006	duo
18963	Prudentius	Hamartigeneia	Praefatio	00002	Quos feminarum prima primos procreat	PruHam00007	Quos
18964	Prudentius	Hamartigeneia	Praefatio	00002	Quos feminarum prima primos procreat	PruHam00008	feminarum
18965	Prudentius	Hamartigeneia	Praefatio	00002	Quos feminarum prima primos procreat	PruHam00009	prima
18966	Prudentius	Hamartigeneia	Praefatio	00002	Quos feminarum prima primos procreat	PruHam00010	primos
18967	Prudentius	Hamartigeneia	Praefatio	00002	Quos feminarum prima primos procreat	PruHam00011	procreat
18968	Prudentius	Hamartigeneia	Praefatio	00003	Sistunt ad aram de laborum fructibus	PruHam00012	Sistunt
18969	Prudentius	Hamartigeneia	Praefatio	00003	Sistunt ad aram de laborum fructibus	PruHam00013	ad
18970	Prudentius	Hamartigeneia	Praefatio	00003	Sistunt ad aram de laborum fructibus	PruHam00014	aram
18971	Prudentius	Hamartigeneia	Praefatio	00003	Sistunt ad aram de laborum fructibus	PruHam00015	de
18972	Prudentius	Hamartigeneia	Praefatio	00003	Sistunt ad aram de laborum fructibus	PruHam00016	laborum
18973	Prudentius	Hamartigeneia	Praefatio	00003	Sistunt ad aram de laborum fructibus	PruHam00017	fructibus
18974	Prudentius	Hamartigeneia	Praefatio	00004	Deo sacranda munerum primordia	PruHam00018	Deo
18975	Prudentius	Hamartigeneia	Praefatio	00004	Deo sacranda munerum primordia	PruHam00019	sacranda
18976	Prudentius	Hamartigeneia	Praefatio	00004	Deo sacranda munerum primordia	PruHam00020	munerum

Image produite par les auteurs · Capture d'écran effectuée sur Excel

44 C'est ensuite dans Heurist que les données intertextuelles ont été complétées à travers un formulaire réalisé selon un modèle scientifiquement établi par l'équipe allemande, que nous avons subdivisé en quatre composantes distinctes : les informations relatives à la source, à la référence ciblée, à l'étude de l'intertextualité et aux métadonnées.

Figure 15. Capture d'écran du premier onglet du formulaire « Intertextuality » dans Heurist

Image produite par les auteurs

45 La création du site de présentation scientifique, qui permet un accès libre aux données du corpus, a été la dernière étape après l'élaboration technique de la base de données. Leur fouille reste malgré tout guidée par

les filtres prédéfinis sur le site, car seuls ceux ayant accès au back-office de la base de données peuvent effectuer des requêtes complètement libres<sup>27</sup> (figure 16).

Figure 16. Recherche des intertextualités chez Paulin de Nole

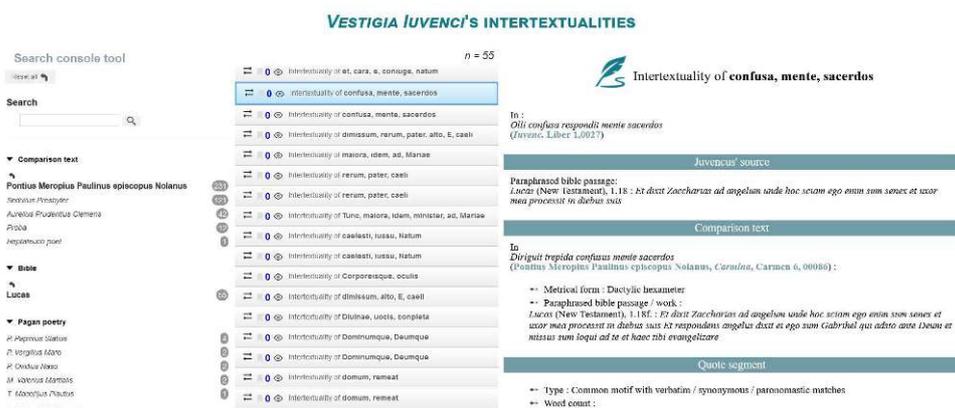


Image produite par les auteurs

46 L'édition numérique hypertexte, qui est la deuxième composante du projet, résulte du croisement entre les fichiers XML-TEI des textes tokénisés et le fichier d'export Heurist en XML-HML de l'intégralité des intertextualités.

47 Techniquement, le format commun XML rend possible l'interopérabilité des deux fichiers, toutefois les deux langages complexifient l'échange des informations car leur grammaire justifie un usage très différent : la TEI est employée pour encoder des textes littéraires dans le cadre de la recherche et le HML pour la visualisation Web. Tandis que le premier permet un encodage riche des textes en arborescence, le deuxième propose une lecture extrêmement linéaire et plate du corpus avec une imbrication hiérarchique en trois niveaux seulement. Pour utiliser ces données et croiser les deux fichiers XML, il fallait donc, soit transformer les données XML-HML en TEI, soit passer par un format passerelle. Nous avons choisi la deuxième option car nous avons besoin de publier les données en HTML et recourir au XML-TEI ajoutait une étape de traitement supplémentaire inutile. Dans cette optique, nous avons eu recours à une feuille de transformation XSLT avec une déclaration préalable des deux grammaires (figure 17) et dont la clé était de faire coïncider les numéros des mots des textes avec ceux faisant l'objet d'une intertextualité dans la base de données.

Figure 17. Déclaration des grammaires HML et TEI dans le fichier XSLT (en jaune)

```
<?xml version="1.0" encoding="UTF-8"?>
<xsl:stylesheet xmlns:xsl="http://www.w3.org/1999/XSL/Transform"
  xmlns:hml="https://heuristnetwork.org"
  xmlns:xs="http://www.w3.org/2001/XMLSchema"
  xmlns:tei="http://www.tei-c.org/ns/1.0"
  exclude-result-prefixes="xs"
  version="2.0">
```

Image produite par les auteurs -Capture d'écran effectuée sur le logiciel Oxygen

48 Dans le fichier XML-HML, tous les formulaires issus de l'exportation sont compris dans une seule balise générique <records> et chacun d'entre eux est renseigné dans un <record> unique qui regroupe une liste de balises sœurs contenant l'intégralité des données. Toutefois, les informa-



différente pour chaque composante. Toutefois, la nécessité d'obtenir un corpus combinant du texte en XML et une structure de base de données nous a permis d'essayer une nouvelle approche, puisqu'il était crucial d'envisager leur gestion d'une toute nouvelle manière pour que les deux éléments puissent communiquer bilatéralement à travers un langage commun. Les données encodées en XML-TEI sont structurées au moyen de balises formant une arborescence, qui offrent une grande latitude par leur variété et permettent une grande précision grâce à leurs attributs. Au contraire, les bases de données relationnelles sont plus rigides dans le sens où elles fonctionnent à travers l'établissement de tables et d'objets préalablement définis. Les exports extrêmement plats XML-HML en sont par ailleurs le reflet. L'utilisation d'Heurist nous a ainsi forcés à anticiper les obligations liées à la double nature du corpus, là où en XML nous avons une forme d'adaptabilité.

51 D'autre part, alors que les transformations XSLT permettent des modifications d'ampleur sans de trop grandes difficultés, la plateforme CMS impose des limites techniques auxquelles il a fallu s'adapter dans la mesure du possible.

## Conclusion

52 L'essentiel de nos réflexions sur les projets numériques de *PoBLAM*, en raison de leur nature spécifique et leurs objectifs divers, s'est focalisé sur le besoin d'uniformité de l'ensemble qui permet un accord équilibré entre les objectifs scientifiques et les pratiques numériques dans le cadre de formats *open source* et interopérables. Nos solutions se sont concentrées sur l'établissement d'un socle stable utilisant les langages XML et HTML, qui ont la robustesse requise sur le très long terme, assurant un maintien et d'une transmission pérenne de l'ensemble des travaux, tout en étant simples d'utilisation pour des chercheurs, y compris novices. Le choix du SGBD Heurist est conforté par le soutien d'Huma-Num et l'assurance de la préservation des données grâce aux différents exports possibles, particulièrement en XML-HML. Ce SGBD, qui est aisément paramétrable à volonté, a contribué à familiariser les chercheurs aux bases de données et à réduire leur appréhension face à la difficulté du travail de saisie. Pour assurer une bonne conservation, il a été décidé dès le début de déposer, pour les trois projets présentés dans cet article, toutes les données numériques, les textes encodés avec les schémas d'encodage et les feuilles de transformation créées, sur l'entrepôt de données Nakala à la fin de l'année 2024. Ces dépôts sont accompagnés de manuels afin d'explicitier chacune de nos décisions techniques. Seul l'accès administrateur de la base de données Heurist *Vestigia Iuenci* est restreint à ses commanditaires, pour des raisons évidentes d'intégrité scientifique.

## Bibliographie

- Boulétreau, Viviane et Benoît Habert. 2014. « Chapitre 9. Les formats ». Dans *Pratiques de l'édition numérique*, édité par Michaël E. Sinatra et Marcello Vitali-Rosati, 145-159. Montréal : Presses de l'université de Montréal. <https://doi.org/10.4000/books.pum.329>.
- Dinkova-Bruun, Greti. 2019. « Marking One's Own : Aegidius of Paris's Revisions of Peter Riga's "Aurora" ». *Filologia Mediolatina* 26 : 127-142.
- Drake, Miriam A., éd. 2003. *Encyclopedia of Library and Information Science*. 4 vol. New York : M. Dekker.
- Manso, José Henrique Rodrigues. 2014. *A edição quinhentista do comentário barbosiano à História Apostólica de Arátor*. Coimbra : Imprensa da Universidade de Coimbra.
- McCarty, Willard. 2003. « Humanities Computing ». Dans *Encyclopedia of Library and Information Science (2<sup>e</sup> édition)*, édité par Miriam Drake, 1224-1235. New York : M. Dekker. <https://www.mccarty.org.uk/essays/McCarty,%20Humanities%20computing.pdf>.
- Pantani, Italo, Margarida Miranda et Henrique Manso. 2014. *Aires Barbosa na cosmópolis renascentista*. Coimbra : Imprensa da Universidade de Coimbra.
- Schreibman, Susan, Ray Siemens et John Unsworth, éd. 2004. *Companion to Digital Humanities*. Oxford : Blackwell Publishing. <http://www.digitalhumanities.org/companion/>.
- Schreibman, Susan, Ray Siemens et John M. Unsworth, éd. 2016. *A New Companion to Digital Humanities*. Malden : Wiley Blackwell.
- Sinatra, Michaël E. et Marcello Vitali-Rosati. 2014. « Chapitre 3. Histoire des humanités numériques ». Dans *Pratiques de l'édition numérique*, édité par Michaël E. Sinatra et Marcello Vitali-Rosati, 49-60. Montréal : Presses de l'université de Montréal. <https://doi.org/10.4000/books.pum.317>.
- Sinatra, Michaël E. et Marcello Vitali-Rosati, éd. 2014. *Pratiques de l'édition numérique*. Montréal : Presses de l'université de Montréal. <https://books.openedition.org/pum/306>.
- Stokes, Peter A. 2020. « eScriptorium : un outil pour la transcription automatique des documents ». *ÉpheNum* (blog). 30 juin. <https://doi.org/10.58079/ofzq>.
- THATCamp Paris. 2011. « Manifeste des *Digital Humanities* ». *THATCamp Paris* (blog). <http://tcp.hypotheses.org/318>.

## Notes

- <https://poblam.u-strasbg.fr>.
- <https://girpam.u-strasbg.fr>.
- Dans ce texte, nous abordons plusieurs *projets* spécifiques inscrits dans le cadre général de *PoBLAM*, qui est un *projet* au sens de l'ANR.
- Pour une étude méthodologique générale, voir les ouvrages de Schreibman, Siemens et Unsworth (2004 ; 2016) et de Sinatra et Vitali-Rosati (2014).
- La transdisciplinarité est un élément-clé de la définition des humanités numériques dans le « Manifeste des *Digital Humanities* » (THATCamp Paris 2011) : « Les *digital humanities* désignent une transdiscipline, porteuse des méthodes, des dispositifs et des perspectives heuristiques liés au numérique dans le domaine des Sciences humaines et sociales ».
- Sur les choix des formats numériques en fonction de leurs usages, voir Boulétreau et Habert 2014.
- Sur les questions épistémologiques des humanités numériques, voir aussi McCarty 2003, 1231-1233.
- « [L'éditorialisation] met l'accent sur les dispositifs technologiques qui déterminent le contexte et l'accessibilité d'un contenu, ainsi que sur la réflexion autour de ces dispositifs ».
- Sur le personnage, voir Pantani, Miranda et Manso 2014 et, sur le commentaire lui-même, Manso 2014.
- Pour l'outil lui-même, voir <https://escriptorium.inria.fr> ; pour une présentation de cet outil, voir Stokes 2020. Les résultats obtenus en phase de test avec Transkribus (<https://www.transkribus.org>) sont également très pertinents ; on aurait pu donc choisir l'un ou l'autre outil.

11 Un cas intéressant pour l'encodage est celui de la désignation périphrastique sur le modèle <persName nymRef="#Augustinus">Aquila hipponensis</persName>.

12 Dans bien des cas cependant, en particulier pour la littérature scolastique, l'identifiant normalisé n'existe pas ou est inventé *ad hoc* par chaque citeur moderne ; cet état de fait, qui devrait normalement se réduire avec la standardisation des références, oblige à imaginer par déduction la forme que pourraient prendre ces références en fonction des pratiques du TLL.

13 <http://hyperdonat.tge-adonis.fr>.

14 <http://poexegese-girpam.u-strasbg.fr>.

15 La *Patrologia Latina* est accessible en ligne sur le site *Corpus Corporum* : <https://mlat.uzh.ch/browser?path=/38/>.

16 Voyant Tools (<https://voyant-tools.org>) est un outil d'analyse de textes numériques, libre d'accès, en ligne et extrêmement simple d'utilisation. En quelques clics, il est possible de sortir des tableaux d'occurrences de lemmes et de leur contexte, des syntagmes, des tendances, des grappes de termes, etc. Son autre avantage est qu'il facilite le partage des résultats sous divers formats et sous différentes formes, qu'ils soient statiques ou dynamiques (PNG, *iframe*, etc.). Nous avons détaillé son utilisation dans le manuel de présentation du projet *Poéxégèse* : <http://poexegese-girpam.u-strasbg.fr/manuels.xhtml>. Voir également le guide officiel : <https://voyant-tools.org/docs/#!/guide>.

17 [http://poexegese-girpam.u-strasbg.fr/analyses\\_textuelles.html](http://poexegese-girpam.u-strasbg.fr/analyses_textuelles.html).

18 Ce projet a donné lieu à un colloque international à Wuppertal en mai 2023, dont les actes paraîtront 2025. Il se situe dans un ensemble de colloques sur la réception des poètes chrétiens, complété par un colloque à Besançon en octobre 2023 sur la réception de Paulin de Nole et de Prudence, dont les actes vont également paraître dans la collection des Études augustiniennes chez Brepols.

19 <https://heuristnetwork.org>.

20 La liste de diffusion [[heurist-utilisateurs@groupe.renater.fr](mailto:heurist-utilisateurs@groupe.renater.fr)] est très réactive et bienveillante ; ses conseils nous ont toujours apporté une aide fructueuse. L'inscription à la liste de diffusion se fait par l'URL <https://groupe.renater.fr/sympa/info/heurist-utilisateurs/>.

21 Nous tenons à remercier Ian Johnson et son équipe pour leur aide et leur soutien indispensables à la réalisation de la base de données.

22 Pour mémoire, HML est le langage XML utilisé pour l'export des données Heurist.

23 Exception faite pour Paulin de Nole : nous avons choisi « PIN » car « Pau » peut prêter à confusion avec « Pru » pour Prudence.

24 Cela ne vaut que pour les auteurs possédant plus d'une seule œuvre. C'est pourquoi nous avons laissé « Pro » pour Proba car il ne pouvait être confondu avec le « Pru » de Prudence, qui est toujours accompagné du nom d'une de ses œuvres.

25 Le premier mot est ainsi numéroté 00001.

26 À l'exception de la ponctuation.

27 L'intégralité des données relatives aux intertextualités se trouve dans le fichier d'export Heurist XML-HML. Des requêtes pourront donc être effectuées depuis ce document et au moyen d'autres logiciels.

## Auteurs

### Bruno Bureau

UMR 5189 HiSoMA, université Jean-Moulin – Lyon III, Lyon, France

Bruno Bureau est professeur de langue et littérature latines, ses recherches portent sur la poésie latine tardive (profane et chrétienne) et sur la littérature de commentaire. Avec Christian Nicolas, il a réalisé l'édition numérique du commentaire des comédies de Térence attribué à Aelius Donat. Il codirige avec Michele Cutino la partie française du projet *Poésie biblique latine antique et médiévale (PoBLAM)* financé par l'Agence nationale de la recherche (ANR) et la Deutsche Forschungsgemeinschaft (DFG) et est en particulier responsable de la réalisation des projets numériques.

ORCID [0000-0001-5558-411X](https://orcid.org/0000-0001-5558-411X)

[bruno.bureau@univ-lyon3.fr](mailto:bruno.bureau@univ-lyon3.fr)

### Inès Burri

UMR 5189 HiSoMA, université Jean-Moulin – Lyon III, Lyon, France

Inès Burri est ingénieure d'études contractuelle chargée de l'élaboration, du développement et de la gestion des techniques et méthodes en humanités numériques dans le cadre du projet *Poésie biblique latine antique et médiévale (PoBLAM)* financé par l'Agence nationale de la recherche (ANR) et la Deutsche Forschungsgemeinschaft (DFG).

ORCID [0009-0009-8821-7768](https://orcid.org/0009-0009-8821-7768)

[ines.burri@univ-lyon3.fr](mailto:ines.burri@univ-lyon3.fr)

## Droits d'auteur



Le texte seul est utilisable sous licence [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/). Les autres éléments (illustrations, fichiers annexes importés) sont « Tous droits réservés », sauf mention contraire.