# Galenus-verbatim, Galien en ligne, un exemple de philologie numérique

Shannon, dans son célèbre [article de 1948](files/test.pdf)[[1]](#footnote-1), a ouvert une nouvelle perspective sur la thermodynamique, en proposant que l’information soit l’envers de l’entropie. Dans ce monde physique, la philologie est alors un effort millénaire contre l’entropie, afin de conserver et ordonner les textes, malgré l’usure de la matière, les accidents de l’histoire, et surtout, l’oubli et l’ignorance. L’informatisation de ce dernier demi-siècle n’est que l’automatisation d’une tradition, les *humanités numériques* sont un humanisme, une imprimerie continuée, avec l’ordinateur. Qui travaille sur des textes anciens se sait héritier et légataire, juste un chaînon dans la transmission. Toute édition, papier ou écran, doit être l’occasion d’apporter au texte. Qu’avons-nous apporté à Galien, qu’est-ce que Galien peut apporter aux humanités numériques ?

[Galenus-verbatim](https://galenus-verbatim.huma-num.fr/)[[2]](#footnote-2) (GV) envisage d’être un site de référence pour les études galéniques. Contrairement à des bases de textes plus générales, comme le TLG (commercial), ou l’excellent Perseus Project, libre depuis plus de 35 ans ; GV prend la responsabilité de satisfaire complètement les spécialistes de l’auteur, afin qu’à terme, la référence électronique (pour les textes libres de droits) puisse faire autorité. Cela signifie que GV reprend sous sa responsabilité des fichiers existants, qu’il en ajoute, et qu’il en assure l’édition continuée, tant pour la correction des mots, que la structure et l’identification des passages. C’est la mission la plus lourde, qui engage sur une durée plus longue que des choix logiciels. Toutefois, cette responsabilité éditoriale doit se matérialiser en ligne.

La durée d’engagement de Galenus-Verbatim oriente les choix technologiques. Ainsi par exemple, pour la relecture des textes, il est difficile de trouver des personnes qui savent bien le grec, il n’est pas raisonnable de leur demander en plus de maîtriser XML/TEI/Epidoc, par conséquent il faut savoir leur adresser les textes dans un format où ils sont rapidement productifs (docx), charge à la technique de réinsérer leur travail dans une structure Epidoc. Il est prudent d’éviter les technologies lourdes (docker, java…) qui demandent un support institutionnel, en essayant de se contenir dans les offres les moins coûteuses (PHP). De manière générale, il faut utiliser des technologies éprouvées, aux limites connues, que les composants soient déboîtables et remplaçables par une alternative, pour s’adapter à des fortunes technologiques très variables sur la durée. Le logiciel de référence pour la publication en ligne de textes antiques, outre le TLG (payant), est [Scaife Viewer](https://scaife.perseus.org/library/urn:cts:greekLit:tlg0057/)[[3]](#footnote-3), du Perseus project. Parce que GV est spécialisé sur un auteur, il fait mieux, avec (beaucoup) moins de moyens. Comment ?

* Une bibliographie ordonnée des traités, hiérarchisant notamment les éditions numérisées et les éditions critiques (production avec [Zotero](https://www.zotero.org/groups/galenus-verbatim/library)[[4]](#footnote-4), publication PHP)
* Une chaîne de numérisation pour ajouter de nouveaux textes, notamment, la traduction latine de Galien dans l’édition de Kühn (20 000 pages). Process : images (OCR ABBYY ou Transkribus) docx (Teinte) XML/TEI (édition XML) Epidoc (Teinte) docx (relecture d’helléniste) docx (Teinte) Epidoc
* Un entrepôt ouvert avec les textes au format Epidoc, conservant l’historique des interventions éditoriales qui pour une part nous précèdent ([GitHub](https://github.com/galenus-verbatim/galenus_cts/)[[5]](#footnote-5)),
* Pour un moteur de recherche lemmatisé du grec ancien et du latin, indexation hors ligne et préparation d’un paquet de publication SQLite (Python, [pip:pie-extended](https://pypi.org/project/pie-extended/)[[6]](#footnote-6) de Thibault Clérice, coordonnés dans [Verbapie](https://github.com/galenus-verbatim/verbapie)[[7]](#footnote-7))
* Des URL pérennes pour le moindre composant textuel : traité/section ou page/ligne/mot (spécification [CTS URN](http://cite-architecture.github.io/ctsurn_spec/)[[8]](#footnote-8), framework léger dédédié aux humanités numériques avec contrôle du routage PHP, [packagist:oeuvres/kit](https://packagist.org/packages/oeuvres/kit)[[9]](#footnote-9))
* Un site qui puisse répondre à ces URL, en affichant les textes et les images libres, dans les limites techniques d’un hébergement mutualisé (PHP-SQlite : [Verbatim](https://github.com/galenus-verbatim/verbatim)[[10]](#footnote-10))
* Un CMS pour les article d’accompagnement, dans le format natif utilisé par les chercheurs, c’est-à-dire le docx des traitements de textes (PHP, [packagist:oeuvres/teinte](https://packagist.org/packages/oeuvres/teinte)[[11]](#footnote-11))

Galenus-Verbatim fonctionne, il peut partir vers une ou deux décennies de vie, avec des pièces réparables, évolutives, ou remplaçables. D’ores et déjà, nous avons apporté aux études galéniques un gain de temps pour la citation de passages de l’œuvre, en assurant notamment la fiabilité des numéros de lignes dans 20 000 pages. Toutefois, à terme, l’enjeu est de ne plus citer l’auteur selon les tomes de l’édition “repère” (Kühn, début XIXe s.), mais relativement à la structure intrinsèque du texte (livre/chapitre/section…). Nous avons donc contribué à lutter contre l’entropie dans Galien.

Tous les développements ont été conçus pour séparer le spécifique galénique de ce qui peut intéresser d’autres projets. Ainsi, il est rapidement possible de monter une édition Epidoc pour d’autres auteurs, et la poser sur un hébergement à quelques euros par an ( ? Cahal). Plusieurs aspects de notre méthodologie, et des composant logiciels qui les implémentent, semblent transposables à d’autres projets éditoriaux : le pilotage bibliographique avec Zotero, la conservation des textes avec un entrepôt git, le CMS word, la normalisation d’un paquet de publication indexé et lemmatisé sous forme de fichier SQLite, un routage totalement contrôlé des URLs pour en garantir la pérennité…

Frédéric Glorieux, juin 2023

1. Shannon, 1948, [*A Mathematical Theory of Communication*](https://people.math.harvard.edu/~ctm/home/text/others/shannon/entropy/entropy.pdf) [↑](#footnote-ref-1)
2. <https://galenus-verbatim.huma-num.fr/> [↑](#footnote-ref-2)
3. <https://scaife.perseus.org/library/urn:cts:greekLit:tlg0057/> [↑](#footnote-ref-3)
4. <https://www.zotero.org/groups/galenus-verbatim/library> [↑](#footnote-ref-4)
5. <https://github.com/galenus-verbatim/galenus_cts/> [↑](#footnote-ref-5)
6. <https://pypi.org/project/pie-extended/> [↑](#footnote-ref-6)
7. <https://github.com/galenus-verbatim/verbapie> [↑](#footnote-ref-7)
8. <http://cite-architecture.github.io/ctsurn_spec/> [↑](#footnote-ref-8)
9. <https://packagist.org/packages/oeuvres/kit> [↑](#footnote-ref-9)
10. <https://github.com/galenus-verbatim/verbatim> [↑](#footnote-ref-10)
11. <https://packagist.org/packages/oeuvres/teinte> [↑](#footnote-ref-11)