

The *Digital Latin Library*

Samuel J. Huskey

University of Oklahoma
huskey@ou.edu

Abstract

This article relates a brief history of the *Digital Latin Library* (DLL), discusses its role in the ever-expanding universe of digital scholarship on Latin language and literature, describes its main projects — the *Library of Digital Latin Texts* (LDLT) and the *DLL Catalog* — and introduces its current research focus on adapting AI/ML technologies to advance its mission.

Keywords: Library; Textual Criticism; Text Encoding; Machine Learning; Digital Scholarly Editions.

Il presente articolo offre una sintetica ricostruzione della storia della Digital Latin Library (DLL), ne esamina il ruolo all'interno dell'universo, in costante espansione, degli studi digitali dedicati alla lingua e alla letteratura latina, descrive i suoi principali progetti — la Library of Digital Latin Texts e il DLL Catalog — e illustra l'attuale indirizzo di ricerca volto ad adattare le tecnologie dell'intelligenza artificiale e dell'apprendimento automatico al fine di promuovere e sviluppare la missione dell'iniziativa.

Parole chiave: Biblioteca; Critica testuale; Codifica del testo; Apprendimento automatico; Edizioni scientifiche digitali.

1. Introduction

The *Digital Latin Library* (DLL)¹ began in 2012 with a planning grant from the Andrew W. Mellon Foundation (now «The Mellon Foundation»)² to the American Philological Association (since renamed as the Society for Classical Studies), in collaboration with the Medieval Academy of

¹ The DLL's main site is <https://digitallatin.org/> (last accessed 21 October 2025). For an earlier overview of the project, see [10].

² <https://www.mellon.org/>, last accessed 21 October 2025. Grant numbers: 11200693, 21400643, 21500706.

America and the Renaissance Society of America.³ The original proposal stated that the admittedly asymptotic goal of the *DLL* was «To create and facilitate use of a virtual, ubiquitous, sustainable working space for scholars and readers of Latin texts of all periods.» By the end of the planning period, the team had adopted a two-fold mission for the project:

- [1] To publish and curate critical editions of Latin texts, of all types, from all eras.
- [2] To facilitate the finding and, where openly available and accessible online, the reading of all texts written in Latin.

After the initial planning period, the Mellon Foundation invited the scholars working on the *DLL* to submit a proposal for a second planning grant. Since I had led the meetings during the planning period, and since the University of Oklahoma (OU), my home institution, had expressed interest in being the long-term host for the project, the team asked me to lead this next stage of the project's development. I recruited two colleagues at OU to be co-principle investigators with me: June Abbas, a professor in the School of Library and Information Studies, and Chris Weaver, a professor in the School of Computer Science.⁴ In 2014 we received the second planning grant, which we used to develop plans for:

- [1] Conducting studies to determine the expectations of different types of user of the library components of the *DLL*;
- [2] Creating tools for visualizing textual data in novel ways to push the boundaries of the concept of «*digital Latin*»;
- [3] Building a platform for publishing born-digital critical editions of Latin texts.

That planning period led to an implementation grant, also from the Mellon Foundation, which began in 2015 and lasted until 2018. By the end of that period, we had delivered several items. We built a digital library catalog according to the outcomes of Abbas' user studies. It consisted of authority files, work records, and the beginnings of a collection of digital versions of Latin texts. Weaver delivered a version of his *Improvise* visualization application that he and his graduate students customized with experimental visualizations of Latin textual data ([18], [19], [20], [21]). I worked with Hugh Cayless, Tom Elliott, Jim Ginther, and several other collaborators to customize the guidelines of the *Text Encoding Initiative* (TEI) for encoding critical editions of Latin texts in Extensible Markup Language [13]. Cayless also developed an interactive web application for viewing XML editions ([5], [6]). We also launched an informational web site and separate web sites for the *Library of Digital Latin Texts* (LDLT) and the *DLL Catalog*.⁵ Finally,

³ The members of the original working group: Roger Bagnall (Columbia University), Hugh Cayless (Duke University), Kathleen Coleman (Harvard), Cynthia Damon (University of Pennsylvania), Tom Elliott (New York University, Institute for the Study of the Ancient World), Michael Gagarin (University of Texas, Austin), Jim Ginther (St. Louis University), Sander Goldberg (University of California, Los Angeles), Jeffrey Henderson (Boston University), Samuel Huskey (University of Oklahoma), Robert Kaster (Princeton), John Miller (University of Virginia), Jim O'Donnell (Georgetown University), Diana Robin (The Newberry Library).

⁴ I would like to dedicate this article to the memory of Chris Weaver, who died unexpectedly in 2025. For more about Weaver, see [23].

⁵ <https://ldlt.digitallatin.org/> and <https://catalog.digitallatin.org/>, both last accessed 21 October 2025.

to encourage the development of a new generation of textual scholars, we published a collection of high-quality videos by experts on editing Latin texts (described below in *The DLL Catalog*).

Since the conclusion of that funding period, the *DLL* project has been housed at the University of Oklahoma, which provides a virtual server and access to institutional space and resources. The *DLL* continues to pursue its two-fold mission through its two main projects: publishing new critical editions through the *LDLT* series, and building a collection of existing texts in the *DLL Catalog*. The rest of this article will describe those projects and introduce the *DLL*'s new focus on applying artificial intelligence and machine learning tools and techniques to accelerate its productivity.

2. *The Library of Digital Latin Texts*

The *LDLT* is a platform for publishing new, born-digital, critical editions of Latin texts from all eras in an open and accessible format. Its main purpose is to solve a long-standing problem with Latin texts on the internet: the lack of any critical apparatus. Although there are some notable exceptions, such as Dániel Kiss' *Catullus Online* and *Musisque Deoque*, most online texts lack not only an *apparatus criticus*, but also any of the other features of traditional critical editions.⁶ Consequently, filling this void has been the main objective of the *LDLT*.

In developing ways of representing the information traditionally included in critical editions, the *LDLT* has broken from the traditional publication paradigm in several ways.⁷ Instead of confining critical editions to the boundaries of the printed page, as they have been for centuries, the *LDLT* treats content and format as separate concerns.

On the one hand, the contents of a critical edition (i.e., the introductory matter, the edition's text, the critical apparatus, etc.) comprise textual data of different, but related, types. In traditional critical editions, the arrangement of these data on the pages of a book constitutes a kind of visual encoding, and the quantity and representation of that information is determined, to a non-trivial extent, by the commercial concerns of the publisher. In *LDLT* editions, however, the top priority is encoding textual data in a way that makes it accessible to both human and machine readers. Since the resulting files are lightweight, platform-independent text files, the quantity of data in an *LDLT* edition is not constrained by commercial concerns about costs and materials.

On the other hand, traditional critical editions are published in a single format: a printed book. Although centuries of development have yielded an efficient way of representing critical edition information on the printed page, nevertheless that format prevents any meaningful computational work from being done with that information. By separating content from its representation, the *LDLT* opens textual criticism to the field of information visualization. The experimental visualizations developed by Weaver and his students for the *DLL* demonstrate how radically our perspective on critical editions can be changed. At the same time, it is important to maintain the connection to more traditional formats, which is why we have also

⁶ <http://catullusonline.org/>, last accessed 21 October 2025; <https://www.mqdq.it/>, last accessed 21 October 2025.

⁷ See [15] for a discussion of the constraints of printed editions, and [17] for criticism. [12] responds to both from a technical perspective.

developed not only an interactive, web-based application for working with *LDLT* editions, but also an XSLT stylesheet for transforming *LDLT* data into the traditional print-based format [9].

Of course, the format of the traditional critical apparatus in print is not entirely the result of commercial concerns. The disciplines of textual criticism and editorial technique have evolved different approaches for different kinds of texts, and those approaches are valid whether editors apply them to print or digital editions. That is why the *LDLT* does not adhere to one particular approach, but rather endeavors to make it possible for editors to use the one that makes the most sense for a given text.⁸

A selection of *LDLT* titles already published or at various stages of production will demonstrate the breadth of approaches this system accommodates:

- Serviani in *Vergili Aeneidos libros IX-XII commentarii* (Robert A. Kaster et al., in progress)
- *Bellum Alexandrinum* (Cynthia Damon, under peer review)
- *P. Ovidii Nasonis Ibis* (Tom Keeline, in progress)
- Paulinus of Venice, *Tractatus de diis gentium et fabulis poetarum* (Roberto Pesce, in progress)
- Pietro d’Abano, *Expositio problematum Aristotelis, Particula XIX: De harmonia* (Caleb Mutch, in progress)
- Peter Gracilis, *Commentary on the Sentences, Book I, Question 1* (Jeff Witt and John Slotemaker, published)
- William of Hedon, *Tractatus de Anima, Distinction 3, chapter 6* (Michael Christensen, published)

The most significant departure from the traditional paradigm is not technical at all, but concerns peer review and publication. Commercial presses recruit scholars to perform peer review of books that other scholars and their institutions will then purchase. The *DLL*, however, has a formal partnership with the three learned societies that participated in its original planning and development. These learned societies, through their research and publications divisions, manage the editorial processes, ensuring that editions meet or exceed professional standards of scholarship. The University of Oklahoma, under the auspices of its Board of Regents, is the official publisher of any editions that receive the imprimatur of a learned society. The *DLL* maintains the technical infrastructure for the publication pipeline and works with scholars to facilitate the encoding of their editions. This partnership ensures that *LDLT* editions undergo rigorous peer review on both textual and technical levels before being published on an open and accessible basis. And since *LDLT* editions are published in distributed version control data repositories, they can be corrected and revised as needed. Our “travelling imprimatur” system ensures that only the version that has been subjected to peer review will be labeled as such [15].

⁸ See [13] for both traditional and digital editorial practices considered outside of publication format.

3. The *DLL* Catalog

The *DLL Catalog* is an ongoing project for finding Latin texts openly available on the internet in any format. Following standard practices of research libraries, we developed a metadata model for the types of information to be stored in the catalog's database, including classes for authority records, work records, and individual expressions of those works as PDFs of printed editions, fully digital editions, HTML pages, or any other format ([1], [2]).

The catalog currently has:

- 3,229 authority records for authors and editors
- 5,261 work records
- 2,470 items

To ensure maximum utility, each authority record contains a canonical name form for the individual and as many variant name forms as we could find in standard reference resources such as the Virtual International Authority File (VIAF).⁹ In addition to human-readable names, we also recorded identifiers such as VIAF ID, Library of Congress ID, Canonical Text Services ID, etc. The same is true for work records. Moreover, every author, work, and item receives a unique *DLL* ID for ease of reference.¹⁰

Currently, authors and works from the Classical period have the most coverage. We are eager to collaborate with other projects to add to the catalog's records for Medieval and Neolatin authors, works, and items.

The *DLL* also hosts a collection of free and openly available videos on editing and textual criticism featuring the following scholars:¹¹

- Stefano Baldassarri, Program Director, ISI Florence: «Criticism of Humanist Texts»
- Greg Crane, Professor of Classics, Winnick Family Chair of Technology and Entrepreneurship Tufts University; Alexander von Humboldt Professor of Digital Humanities: «Textual Criticism, Past and Present»
- Cynthia Damon, Professor of Classical Studies, University of Pennsylvania: «Reading the Critical Apparatus»
- Greta Franzini, Postdoctoral Researcher, Göttingen Centre for Digital Humanities: «Digital Editing and Textual Criticism»
- Robert Kaster, Kennedy Foundation Professor of Latin Language and Literature, Princeton University: «Constructing the Critical Apparatus»

⁹ <https://viaf.org/>, last accessed on 21 October 2025.

¹⁰ Alison Babeu, Digital Librarian and research coordinator for the *Persens* project (<https://www.perseus.tufts.edu/hopper/>, last accessed 23 October 2025), generously shared data from the *Persens* Catalog (<https://catalog.perseus.org/>, last accessed 23 October 2025) during this stage of the project.

¹¹ <https://digitallatin.org/videos>, last accessed 21 October 2025.

- James C. McKeown, Professor of Classics, University of Wisconsin: «Textual Criticism»
- Robert H. Rodgers, Lyman-Roberts Professor of Classics, University of Vermont: «Introduction to Critical Editing»
- Richard Tarrant, Pope Professor of the Latin Language and Literature, Harvard University: «Interpolation»

4. The *DLL* and Artificial Intelligence

From the beginning of the project the “digital” in “*Digital Latin Library*” has referred not only to the digital format of texts in the *LDLT* and the *DLL Catalog*, but also to the development and use of digital tools and methods for working with information. We have built tools for extracting information from websites and online catalogs and manipulating it for various purposes.¹² We have also created tools and workflows for automating the tedious task of encoding critical editions in XML according to our guidelines [8]. In recent years, as the field of Artificial Intelligence and Machine Learning (AI/ML) has become more accessible, we have looked increasingly toward AI to facilitate the *DLL*’s mission.

For example, thousands of records of Latin texts can be found and downloaded from a research database such as the *HathiTrust Digital Library* in a matter of minutes.¹³ Reconciling one record to the corresponding authority and work records in the *DLL Catalog* is the work of only minutes, but reconciling thousands of records can consume weeks or even months, depending on the variety of name forms and titles in the records. That is because it is easy to match *P. Vergili Maronis Aeneis* with the authority record for Virgil and the work record for *Aeneis*, but it is much more difficult to decide what to do with a title such as the following :¹⁴

Hoc uolumine continentur poëtae tres egregij nunc primum in lucem editi:
Gratij, q[ui] Augusto Principe floruit, de uenatione lib. I. P. Ouidij
Nasonis Halieuticôn liber acephalus. M. Aurelij Olympij Nemesiani
Cynegeticôn lib. I ; eiusdem carmen bucolicum. T. Calphurnij Siculi
Bucolica. Adriani Cardinalis uenatio.

One must also reckon with the dozens of variant forms of author names and work titles that different catalogers in different countries have assigned to different books and editions. To assist with the tedious, time-consuming task of working with noisy bibliographic records, we have fine-tuned a version of the Bidirectional Encoder Representations from Transformers (BERT)¹⁵ model to rapidly and accurately match variant names of authors in results from external catalogs

¹² Many of these tools are available in the *DLL*’s GitHub repositories at <https://github.com/DigitalLatin>, last accessed 21 October 2025. On the *DLL*’s use of Git itself as a research tool, see [10].

¹³ <https://www.hathitrust.org/>, last accessed 21 October 2025.

¹⁴ This is how the title of an Aldine edition published in 1534 appears in the *HathiTrust Digital Library* (<https://hdl.handle.net/2027/uc1.31175035491763>, last accessed 21 October 2025).

¹⁵ Specifically, I used the DistilBERT base multilingual (cased) model, described in [17].

with their identifiers in the *DLL Catalog*.¹⁶ We are also currently refining a script that uses LatinCy [4] pipelines to match variant title forms to their *DLL* identifiers. Yet another project under development uses computer vision models to identify the contents of critical editions in PDF format.

We also have plans to use machine learning to create a concordance tool for using the critical apparatus from multiple editions side by side, potentially as a supplement for *LDLT* editions. This is in response to the suggestion of Cynthia Damon [7], who has been an advisor to the *DLL* since its inception:

I would love to be able to extract readings from the subset of apparatus notes that concern variants, of course, but if I had to choose between recording the readings of five witnesses and recording the notes of five apparatuses I would choose the latter every time. (p. 217)

5. Conclusion

While the *DLL* remains focused on its mission of publishing new critical editions of Latin texts and facilitating the finding and use of digital Latin texts in general, it is also imperative to consider the future of the project. Although institutional support remains strong, external sources of funding have dwindled in recent years, with the consequence that the *DLL* has operated for some time with a skeleton crew of one, albeit with some occasional collaboration from volunteers. Fortunately, many other individuals and teams are also working to expand research and scholarship on the Latin language and literature further into the digital frontier, as is clear from the presentations at the *Digital Latin II* international workshop and the other papers collected in this volume. Moreover, developments in artificial intelligence and machine learning have great potential for multiplying and enhancing those efforts. It seems appropriate, therefore, to conclude this piece with a call to an even greater emphasis on openness, interoperability, accessibility, and, above all, collaboration among the scholars and researchers working in this space. May there be more occasions like the *Digital Latin II* International Workshop!

Bibliography

- [1] Abbas, June, Stacey Renee Baker, Samuel J. Huskey, and Chris Weaver. 2015. “*Digital Latin Library*: Information Work Practices of Classics Scholars, Graduate Students, and Teachers”. *Proceedings of the Association for Information Science and Technology* 52 (1): 1-4. <https://doi.org/10.1002/pr2.2015.1450520100117>.
- [2] Abbas, June, Stacey Renee Baker, Samuel J. Huskey, and Chris Weaver. 2016. “How I Learned to Love Classical Studies: Information Representation Design of the *Digital Latin Library*”. *Proceedings of the Association for Information Science and Technology* 53 (1): 1-10. <https://doi.org/10.1002/pr2.2016.14505301045>.

¹⁶ An article on this project is forthcoming in the *International Journal of Arts and Humanities Computing*.

- [3] Asokarajan, Bharathi. 2016. “A Pixel-Based Focus+Context Technique for Visualizing Variation in Classical Text”. Master’s thesis. University of Oklahoma. <https://shareok.org/handle/11244/34669>.
- [4] Burns, Patrick J. 2023. “LatinCy: Synthetic Trained Pipelines for Latin NLP”. arXiv:2305.04365. Preprint, arXiv. <https://doi.org/10.48550/arXiv.2305.04365>.
- [5] Cayless, Hugh. 2018. “Critical Editions and the Data Model as Interface”. In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider. 249-263. Herstellung und Verlag: Books on Demand GmbH.
- [6] Cayless, Hugh, and Raffaele Vigiante. 2018. “CETEIcean: TEI in the Browser”. Paper presented at Balisage: The Markup Conference 2018, Washington, DC. <https://doi.org/10.4242/BalisageVol21.Cayless01>.
- [7] Damon, Cynthia. 2016. “Beyond Variants: Some Digital Desiderata for the Critical Apparatus of Ancient Greek and Latin Texts”. In *Digital Scholarly Editing*, edited by Matthew James Driscoll and Elena Pierazzo. 201-218. Open Book Publishers. <https://www.jstor.org/stable/j.ctt1fzhh6v.15>.
- [8] Felkner, Katy, and Samuel J. Huskey. 2019. “DigitalLatin/Automation: First Release”. V. v0.1.0. Zenodo. <https://doi.org/10.5281/zenodo.2658751>.
- [9] Huskey, Samuel. 2022. “The Visual [Re]Presentation of Textual Data in Traditional and Digital Critical Editions”. *Magazén* 3 (1): 115-138. <https://doi.org/10.30687/mag/2724-3923/2022/05/005>.
- [10] Huskey, Samuel J. 2019. “The *Digital Latin Library*: Cataloging and Publishing Critical Editions of Latin Texts”. In *Digital Classical Philology: Ancient Greek and Latin in the Digital Revolution*, edited by Monica Berti. 19-34. De Gruyter Saur. <https://doi.org/10.1515/9783110599572-003>.
- [11] Huskey, Samuel J. 2024. “Committing to Reproducibility and Explainability: Using Git as a Research Journal”. *International Journal of Digital Humanities* 6 (1): 9-21. <https://doi.org/10.1007/s42803-023-00076-9>.
- [12] Huskey, Samuel J., and Hugh Cayless. 2022. “The Digital Critical Apparatus: Thoughts from the Field”. *The Classical Journal* 117 (3): 340-58.
- [13] Huskey, Samuel J., Hugh A. Cayless, and Andrew Dunning. 2020. “DigitalLatin/Guidelines: Version 1.3.0”. Zenodo. <https://doi.org/10.5281/zenodo.3698122>.
- [14] Huskey, Samuel J., and Robert A. Kaster. 2024. “Editing”. In *The Cambridge Critical Guide to Latin Literature*. Edited by Roy K. Gibson and Christopher Whitton. 516-562. Cambridge University Press.
- [15] Huskey, Samuel J., and Jeffrey C. Witt. 2019. “Decoupling Quality Control and Publication: The *Digital Latin Library* and the Traveling Imprimatur”. *Digital Humanities Quarterly* 13 (4).

-
- [16] Keeline, Tom. 2017. "The Apparatus Criticus in the Digital Age". *The Classical Journal* 112 (3): 342-363. JSTOR. <https://doi.org/10.5184/classicalj.112.3.0342>.
- [17] Olson, S. Douglas. 2019. "Further Notes on the Apparatus Criticus". *The Classical Journal* 114 (3): 330-344. JSTOR. <https://doi.org/10.5184/classicalj.114.3.0330>.
- [18] Sanh, Victor, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2020. "DistilBERT, a Distilled Version of BERT: Smaller, Faster, Cheaper and Lighter". arXiv:1910.01108. Preprint, arXiv. <https://doi.org/10.48550/arXiv.1910.01108>.
- [19] Silvia, Shejuti. 2016. "VariantFlow: Interactive Storyline Visualization Using Force Directed Layout". Master's thesis. University of Oklahoma. <https://shareok.org/handle/11244/44937>.
- [20] Sunchu, Vamshi Krishna. 2016. "A Flexible Schema-Aware Mapping of XML Data into Relational Models". Master's thesis. University of Oklahoma. <https://shareok.org/handle/11244/44903>.
- [21] Vangala, Sudarshan Reddy. 2016. "Recording and Replaying User Intentions in Coordinated Multiple View Visualizations". Master's thesis. University of Oklahoma. <http://hdl.handle.net/11244/44904>.
- [22] Witt, Jeffrey C. 2018. "Digital Scholarly Editions and API Consuming Applications". In *Digital Scholarly Editions as Interfaces*, edited by Roman Bleier, Martina Bürgermeister, Helmut W. Klug, Frederike Neuber, and Gerlinde Schneider. 219-247. Herstellung und Verlag: Books on Demand GmbH. <http://www.uni-koeln.de/>.
- [23] Wong, Pak Chung, et al. 2025. "Gone Too Soon, Remembered Always: Chris Weaver and the Power of Visual Thinking". *IEEE Computer Graphics and Applications* 45 (5): 8-11. <https://doi.org/10.1109/MCG.2025.3598531>.