

Proteus: A Platform for Born Digital Critical Editions of Literary and Subliterary Papyri

Alex C. Williams^{*†§}, Annapaola Santarsiero^{*}, Chiara Meccariello^{*}, Gertjan Verhasselt^{‡*},
Hyrum D. Carroll[†], John F. Wallin[†], Dirk Obbink^{*}, and James H. Brusuelas^{*}

^{*}University of Oxford, Oxford, UK

Email: {Annapaola.Santarsiero, Chiara.Meccariello, Dirk.Obbink, James.Brusuelas}@classics.ox.ac.uk

[†]Middle Tennessee State University, Murfreesboro, TN, USA

Email: {Alex.Williams, Hyrum.Carroll, John.Wallin}@mtsu.edu

[‡]KU Leuven, Leuven, Belgium

Email: {Gertjan.Verhasselt}@arts.kuleuven.be

[§]University of Waterloo, Waterloo, ON, CA

Abstract—Scholars working on vast archives of ancient fragmentary manuscripts have long faced the problem of managing and intuitively interfacing the exponential growth of data as fragments are re-edited over time. Multiple readings, multiple versions, and the reality of a multi-text are concepts not well-suited to the physical codex. Current digital archives and libraries for ancient Greek and Latin texts provide only a single reading of a text. Furthermore, the current information model for these digital texts fails to include the vital components necessary to create complete born digital critical editions which are citable and usable in Classics. Collectively, these problems have undermined the effectiveness and reputability of digital editions on the World Wide Web. In this paper, we introduce Proteus, a platform designed to facilitate the creation and scholarly use of complete born digital critical editions of literary and subliterary papyri.

Keywords—Digital philology, born digital critical editions, papyrology, text markup, collaborative editing

I. INTRODUCTION

A substantial part of our knowledge of ancient Greek literature is dependent on papyri. Greek works in prose and verse that had not been known for over a millennium have surfaced mostly from the sands of Egypt, crucially complementing the canonical works preserved through the channel of the medieval scribal transmission.

As the dedicated discipline of papyrology evolves and benefits from the application of new imaging technologies, with new fragments being edited and known fragments being re-edited multiple times, the papyrologist’s dataset continually increases and changes over time. A digital environment is the ideal place to interface this mutability and the ongoing scholarly discourse around it as it is capable of encompassing and managing this growing wealth of data in one place rather than an ever increasing number of printed books and journal articles.

Current digital platforms for ancient Greek and Latin texts provide only a single reading of a text, thus reproducing the standard reading experience of the physical codex. Furthermore, the current information model for Greek and Latin digital texts fails to include the vital components necessary to create complete born digital critical editions and facilitate the scholarly use and citation of such editions. In the case of

papyri, these components include an articulated text, a critical apparatus, a diplomatic text, a paleographical apparatus, text translations, images, and metadata. In addition, specific typologies of fragments require a testimonia apparatus and a transcription of ancient marginalia.

In this paper, we introduce Proteus¹, a platform designed to facilitate the creation and scholarly use of complete born digital critical editions of literary and subliterary Greek papyri. Intended to be a first place for publication of new editions of previously published and unpublished material, Proteus specifically focuses on papyri preserving works that have not been handed down to us through the medieval transmission, and for which the papyri are our exclusive or primary textual source. The platform’s name derives from the shared mutability of the mythological figure of the same name, who shifted shape frequently to avoid capture, and critical editions of literary and paraliterary papyri, which change form as they are re-edited over time.

Proteus leverages the expertise of papyrologists and classicists from all over the world to build a dynamic ecosystem for the creation of born digital critical editions of fragmentary texts, proper philological editions as citable and usable in scholarship as those published in established journals and printed books. While embracing the practice of critical editing and setting a new standard for next-generation editions not only of literary papyri, but of ancient texts in general, Proteus also introduces functionalities of which the codex format is incapable, and leverages computing to improve the reading experience.

II. RELATED WORK

Currently, the largest online datasource for papyrological material is Papyri.info², a web application that aggregates information from a number of online databases containing documentary papyrological material, such as court records, tax forms, and other miscellaneous material written on papyrus and ostraca. The application is made-up of two components: the Papyrological Navigator provides capabilities to search, browse,

¹<http://www.proteusproject.uk/>

²<http://www.papyri.info/>

and aggregate papyrological material; and the Papyrological Editor, a variant of the Son of Suda Online (SoSOL) Editor [1], allows users to contribute texts, translations, commentary, and related metadata. Despite being a focal point for modern digital papyrology, the application targets only documentary papyri and consequently cannot be used to create born digital critical editions of literary papyri. Although plans have been announced to extend its functionality to literary papyri³, we are unable to evaluate their proposed system as the application's changes are still a work-in-progress. While the project is unrelated to the task of creating born digital critical editions, the Perseids project [2], which also utilizes the SoSOL editor, should be noted for its goal in acting as collaborative platform for annotating text re-uses of ancient lost works within existing texts in the Perseus database of Greek literary works [3].

To date, there are only three research efforts in classics that have aimed to establish born digital editions of ancient literary texts. The first is Catullus Online⁴, a critical edition of Catullus' poems which includes the text, a critical apparatus containing all modern conjectures on the poems, an overview of ancient authors quoting them (testimonia), and photographs of some of the manuscripts. Though encouraging users to notify the appearance of new conjectures, the project is not intended as the first place for their publication, but presupposes traditional publication in formal peer-review journals. Moreover, due to the specific dynamics of Catullus' transmission, the platform does not aim at editing papyri.

The second research effort is the Homer Multitext project⁵ [4], a critical framework for exploring the multiformity of the text of the Homeric *Iliad* and *Odyssey*. To this purpose, the project provides diplomatic transcriptions of the text and marginalia as preserved in each manuscript, and allows users to contribute in the form of diplomatic transcripts, images, and translations. Since the specific purpose of the project is to capture the state of the text in each of its witnesses, the Homer Multitext does not include the other components required to meet the defined criteria for a complete born digital critical edition, most notably the critical apparatus.

The third research effort is the Catalogue of Paraliterary Papyri⁶, a digital archive of paraliterary Greek papyri. The project intends to complement the Papyri.info database, which focuses exclusively on documentary papyri, by cataloguing a specific category of subliterate or quasi-literary papyri. For each text, it offers an articulated text, a critical apparatus, and relevant metadata. The project does not leverage or encourage collaborative editing as each edition has been manually entered into the system's database by project administrators. Additionally, the catalogue has not been updated since 2007.

While each is useful in the context of its own research goals, these projects do not support the creation of complete born digital critical editions of literary or subliterate papyri.

III. PROTEUS ARCHITECTURE

The Proteus architecture consists of two major components: the Proteus Search Interface and the Digital Editor for Clas-

sical Philology (DELPHI). The project is implemented using Python, HTML5, CSS, JavaScript, the PostgreSQL database management system, and Apache Solr for fast, powerful search functionality. These components are packaged together using Django, a high-level Web framework for the Python programming language. A full description of the components and standards presented in this section can be found online at <http://www.proteusproject.uk/>.

A. Markup Standards for Digital Philology

1) *CSYN-P: An XML Schema for Papyri*: For more than a decade, the EpiDoc standard [5] has been used in most, if not all, online research efforts in digital papyrology. As the standard was designed for epigraphy, many of the standard's XML tags and attributes provide little to no meaning in the context of papyrology and obfuscate the XML structure of a literary papyrological edition. In order to simplify the task of creating XML editions and improve the readability of XML documents, we have designed a new XML standard for philological studies of papyrological material, named the Critical Syntax for Papyri, or CSYN-P. The standard is a simplified and updated implementation of the EpiDoc standard that has been modified to satisfy the requirements for capturing the semantic and structural elements of literary and subliterate manuscripts.

2) *CSYN Markdown*: Studies suggest that using a markup language with a plain text formatting syntax in lieu of a more complex data format, such as XML, can yield higher data quality, improve accessibility, simplify the task of data entry, and hasten the overall data collection process [6]. One such example of a markup language is the Leiden+ standard⁷, which was developed as part of the Integrating Digital Papyrology project to bring ease to the task of entering editions in Papyri.info [7]. In order to reduce, or even remove, the requirement for users to memorize the CSYN-P standard, we have designed and implemented CSYN Markdown, a new, human-readable markup language for digital philology, into DELPHI (see Fig. 1). The syntax of the new markup language, which bears a resemblance to the Leiden+ standard, was created by combining standards from the Leiden Conventions and the popular Markdown language [6] used by many professional and end users. Major differences in comparison to the Leiden+ standard include functionality for denoting the structure of columns and fragments, reducing the number of arbitrary characters or symbols that are irrelevant to the edition, and generally improving readability and intuitiveness of the syntax.

Although CSYN offers a variety of benefits for productivity and accessibility, Proteus does not require a user to edit a text using the markup language. During the editing process, a user can edit a text in either CSYN-P, CSYN Markdown, or toggle between the two formats. By allowing users to choose a method of data entry, we ensure that they can leverage their expertise as a classicist or computer scientist as they see fit.

B. Proteus Search Interface

1) *A Simple, Intuitive Interface*: Modern research in interface design and usability has suggested that many search

³<http://isaw.nyu.edu/news/digital-literary-papyri/>

⁴<http://catullusonline.org/>

⁵<http://www.homermultitext.org/>

⁶<http://cpp.arts.kuleuven.be/>

⁷<https://wiki.digitalclassicist.org/Leiden-plus>

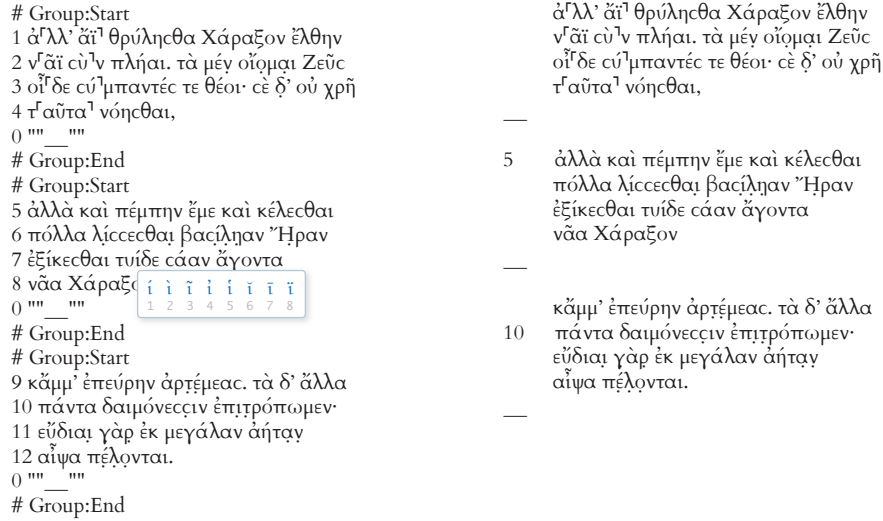


Fig. 1. A snippet of Sappho's "Brothers Poem" in CSYN Markdown (left) and its rendered text (right) in DELPHI.

interfaces are difficult to comprehend because their usage requires knowledge about the underlying data being searched [8]. This problem is particularly true for search interfaces designed for the niche discipline of papyrology. In order to improve the usability of the system for professionals both new and familiar to papyrology, Proteus captures search queries using a single input field where users can query for authors, titles, editors, and content. Once submitted, the query is processed through a faceted search, which allows the user to continuously filter and sort the retrieved results by every possible combination of edition labels (*e.g.*, 1st century) or edition attributes (*e.g.*, critical editors). Studies have suggested that a faceted search approach provides more information-seeking support to search applications than traditional best-first approaches [9][10].

2) *Rendering a Critical Edition without XSLT*: XSLT, a technology for transforming and translating XML documents to different formats (*e.g.*, HTML), has largely become the standard for rendering critical editions online. Despite the technology's overwhelming popularity, many search engines (*e.g.*, Google) do not index XSLT-generated web pages. Furthermore, XSLT stylesheets, which define how the XML document should be transformed, can be difficult to both design and maintain. In order to avoid these complications, Proteus extracts document information using a custom XML parser for the CSYN-P standard and renders the extracted information using Django's internal templating language. An additional benefit of our approach to document processing is that we chose to implement the parser in an imperative programming language (Python), which studies suggest to be more digestible and less confusing for developers that are unfamiliar with functional scripting languages, such as XSLT [11].

C. Digital Editor for Classical Philology

1) *In-Browser Greek Keyboard and Accent Menu*: A key issue in creating born digital editions is standardizing keyboard input. For the past decade, digital classicists have been required to download, install, and purchase a software product

that maps their non-Greek keyboard to the Greek alphabet (*i.e.*, GreekKeys⁸, MultiKey⁹) in order to contribute to online epigraphical or papyrological research efforts. To overcome this requirement, a robust, portable Greek keyboard mapping tool was implemented into DELPHI where every uppercase and lowercase alphabetic ASCII key code is mapped to a Greek letter. With this tool, users have the ability to toggle Greek typing as needed without downloading any additional software. Using DELPHI's button menu, users can insert papyrological characters (*e.g.*, early Christian symbols) that cannot be intuitively mapped to the standard keyboard. Inspired by the Apple OS X Character Accent Menu, a built-in menu for accents and other diacritics was implemented into DELPHI and can be activated by holding the ALT key when the cursor is positioned after a valid letter (see dialog box in Fig. 1).

2) *Real-Time Preview of Editions*: In order to help users adapt to this style of editing and further improve the quality of their contributions, DELPHI visualizes and previews an edition in real-time using the same XML parser described in Section III.B.2. By default, DELPHI is viewed in a dual-pane layout with the editing field on the left and the visualization field on the right (see Fig. 1). After a key has been pressed, DELPHI will automatically translate the user's contributions written in CSYN Markdown or CYSN-P into HTML and render the edition as it would appear in the Proteus database.

An additional benefit in automatically translating a user's submission is that syntactical errors in both CSYN Markdown and CYSN-P, which cause the parser's translation to fail or halt, are revealed to the user in real-time. If a user makes a syntactical error, the visualization field will instantly present an explicit, verbose error message containing the line number, character position, and respective character that produced the error. The traditional transformation approach that leverages XSLT stylesheets does not present any notification of failure in completing the transformation due to a syntactical error. As

⁸<http://apagreekkeys.org/>

⁹<http://www.oeaw.ac.at/kal/multikey/>

a result, this feature reinforces the usefulness and practicality of an XSLT-less transformation approach.

IV. PROMOTING SCHOLARLY COLLABORATION AND USE

A. *Crafting Born Digital Critical Editions of Literary Papyri*

Proteus employs a new collaborative editing model that facilitates parallelism in the editing process. To create a new edition of a text, a user first submits a request that includes a relevant fragment, an explanation as to why the request is being made, and a selection of edition sections on which he wishes to work. At the source level, Proteus enforces that only one user may work on one section of a text at a time. For example, no two users can work on the testimonia of a text simultaneously. However, multiple users can work on unique sections of a text at the same time. By enforcing contributions in this manner, Proteus can easily record how each user contributed to an edition.

Once an administrator has approved a request to contribute, the user can edit the text appropriately and submit his contributions. For each submission, the Proteus Review Board, a peer-review committee consisting of papyrologists from institutions from around the world, can accept, deny, or recommend revision for a submission. If the committee votes for approval, the submission is instantaneously added to Proteus's publicly searchable knowledgebase and subject to criticism from other users. If the committee votes for denial, the submission is permanently deleted. If the committee recommends revision, the submission is returned to the user, who then has the option to re-submit his revised contributions.

To incentivize scholarly contribution, Proteus publicly acknowledges users for their contributions. If a user has created or added to a born digital critical edition, he will be listed as a Proteus Critical Editor for the edition. If the user is simply entering an edition of a text that has been published in print, he will be listed as a Proteus Contributor. If a user has provided a translation for a text, he will be listed as a Proteus Translator.

B. *Facilitating an Ecosystem for Digital Philology*

Scholarly debate is vital to philology. Once an edition has an approved articulated text and critical apparatus, users may offer editorial criticism by adding a critical note to the edition. Like contributions for the sections of the editions themselves, both the request and the submitted critical note must be approved by the Proteus Review board.

For each text in Proteus, users can select from a dropdown box which lists and enumerates the available editions of a document. Upon clicking a edition number in the dropdown list, a dual-pane display will overlay the Proteus Search Interface and render two unique editions simultaneously. This dual-edition perspective both simplifies and hastens the task of comparing and criticizing multiple editions of a text simultaneously.

V. CONCLUSION

Despite digital papyrology having been given rigorous attention in classical research, no computational platform currently exists for intuitively managing the exponential growth of data as fragments are re-edited over time. In this paper, we introduced Proteus, a platform for digital born critical editions

of literary and subliterate Greek papyri. Proteus consists of two major architectural components, the Proteus Search Interface and Digital Editor for Classical Philology, that were designed and engineered with a focus on usability and interaction. As a research platform, Proteus redefines how the modern papyrologist engages ancient texts and encourages a new model and mentality beyond the physical codex.

A. *Future Work*

A key goal for future work is extending the project's functionality to create born digital critical editions of non-papyrological material. For each type of such material, a new subset of the CSYN standard will be designed (*e.g.*, CSYN-M, a critical syntax for medieval manuscripts). Although the XML schema will vary by material, the CSYN Markdown standard will translate to each XML standard appropriately and can still be used to mark up any type of document.

ACKNOWLEDGMENTS

The authors thank the Arts and Humanities Research Council for providing funding for this project. They would also like to acknowledge the Egypt Exploration Society for providing access to the Oxyrhynchus papyri collection.

REFERENCES

- [1] R. Baumann, "The Son of Suda On-Line," in *The Digital Classicist 2013*, S. Dunn and S. Mahony, Eds. London: The Institute of Classical Studies University of London, 2013.
- [2] B. Almas and M. Berti, "Perseids collaborative platform for annotating text re-uses of fragmentary authors," in *Proceedings of the 1st International Workshop on Collaborative Annotations in Shared Environment: Metadata, vocabularies and techniques in the Digital Humanities*. ACM, 2013, p. 7.
- [3] D. A. Smith, J. A. Rydberg-Cox, and G. R. Crane, "The Perseus Project: A digital library for the humanities," *Literary and Linguistic Computing*, vol. 15, no. 1, pp. 15–25, 2000.
- [4] G. Nagy, "Homer Multitext project," *Online Humanities Scholarship: The Shape of Things to Come*, 2010.
- [5] G. Bodard, "EpiDoc: Epigraphic documents in XML for publication and interchange," *Latin on Stone: Epigraphic Research and Electronic Archives, Roman Studies: Interdisciplinary Approaches*, 2009.
- [6] J. Voegler, J. Borschein, and G. Weber, "Markdown—a simple syntax for transcription of accessible study materials," in *Computers Helping People with Special Needs*. Springer, 2014, pp. 545–548.
- [7] R. Bagnall, "Integrating digital papyrology," *Online Humanities Scholarship: The Shape of Things to Come*. <http://hdl.handle.net/2451/29592>, 2010.
- [8] M. Hearst, *Search User Interfaces*. Cambridge University Press, 2009.
- [9] B. Kules, R. Capra, M. Banta, and T. Sierra, "What do exploratory searchers look at in a faceted search interface?" in *Proceedings of the 9th ACM/IEEE-CS joint conference on Digital libraries*. ACM, 2009, pp. 313–322.
- [10] K.-P. Yee, K. Swearingen, K. Li, and M. Hearst, "Faceted metadata for image search and browsing," in *Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, 2003, pp. 401–408.
- [11] V. Tirronen, "Study on difficulties and misconceptions with modern type systems," in *Proceedings of the 2014 conference on Innovation & technology in computer science education*. ACM, 2014, pp. 303–308.