

Enhancing South Slavic Cyrillic Manuscripts Research Through a Digital Toolkit for Cyrillic Palaeography

Marta Riparante¹, Maxim Goynov², Desislava Paneva-Marinova², Lilia Pavlova²

¹ Cyrillo-Methodian Research Center, Bulgarian Academy of Sciences, Bulgaria

² Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Bulgaria

Abstract

This paper presents an innovative digital tool designed to support the analysis of medieval Cyrillic handwriting within a codicological and palaeographical framework. The Cyrillic Palaeography Toolkit (CyPaT), developed from a standardised descriptive model and integrated with the Repertorium of South Slavonic Manuscripts and Copyists (X–XIV cc.), offers a comprehensive environment for the description, processing, and comparative study of manuscript data. By enabling direct interaction with digitised images, CyPaT allows for detailed examination of script features such as letter proportions, stroke composition, and layout. As a free and open-source resource, it promotes collaborative research and the systematic identification of script features linked to specific periods, scriptoria, or individual scribes, particularly within the South Slavic Cyrillic tradition, marking a significant step forward in the development of Digital Slavic Palaeography as an emerging field of scholarly inquiry.

Keywords:

Digital Toolkit, Cyrillic Palaeography, South Slavic Manuscripts, Digital Humanities.

CCS Concepts

• **Applied computing** → Arts and humanities; • **Applied computing** → Document management and text processing; **Information systems** → Information retrieval → Specialized information retrieval; • **Information systems** → Information systems applications → Digital libraries and archives

1. Introduction

Between the 9th and 14th centuries, the Cyrillic alphabet was instrumental in shaping the linguistic and cultural identity of Slavic societies, serving as the foundation for their written traditions. The preserved manuscript heritage from this period offers crucial insights into the evolution of Slavic writing, making palaeographic analysis a central scholarly pursuit. Tracing the formation and transformation of letterforms across time, regions, and scribal contexts is essential to this endeavour; however, the complexity and volume of such analysis increasingly expose the limitations of traditional methods. Recent case study re-evaluations [RD23], and advances in software for cultural heritage management [P-MGZ*23], underscore the growing necessity for digital tools capable of supporting accurate script visualisation, systematic data organisation, and comparative manuscript analysis. Within this context, the present research proposes an integrated digital framework for the study of medieval Cyrillic manuscripts from the South Slavonic tradition. Section 2 introduces the *Repertorium of South Slavonic Manuscripts and Copyists (X–XIV cc.)* [Rep25], as a structured research database, while Section 3 presents the *Cyrillic Palaeography Toolkit* (hereafter CyPaT) [Cyr25], an open-source application designed for the detailed examination of script features directly from digitised images. This section outlines the technological foundation and key functionalities of CyPaT, including image calibration, proportional analysis, stroke-based letter reconstruction, and pattern matching, which together support

a data-driven approach that brings traditional palaeography into dialogue with digital analysis.

2. The Repertorium: mapping Cyrillic manuscripts and script

This study's interpretative model of the Cyrillic script's evolution seeks to reveal key features and graphic transformations through stroke-by-stroke tracing. This approach is supported by the Repertorium, an online platform serving as both a comprehensive database and research tool focused on the script's development between the 10th century and the 14th-century golden age of South Slavic manuscript production. This initiative is part of the research project *Development of the Cyrillic alphabet from the 9th to the 14th century in the South Slavic lands: research and digital presentation* and evolved from an earlier database created within the framework of the research project *Fourteenth Century South Slavonic Scribes and Scriptoria (palaeographic attribution and online repertorium)* [Res25].

The platform standardises a range of distinct entities, including scriptoria, copyists, types of script, and manuscript typologies, with the core unit being the 'Manuscript' object. Each manuscript is described through a model structured across three levels: 'Identification', 'Description', and 'Palaeography' [P-MSR*22] [LGP-M*24]. The most detailed level, 'Palaeography', focuses on the copyist and his handwriting, starting with the traditional classification of the script, drawn from existing catalogues, and including data on the scribe's identity, notes, and manuscript images. This level is further structured into three sublevels: the first

analyses the general appearance of the script, establishing its key proportions, writing angles, and text inclination; the second undertakes a detailed stroke-by-stroke examination of selected letterforms; and the third addresses the modifications applied to these letters, along with the punctuation marks and diacritics employed [Rip24]. The need to support this detailed analysis led to the creation of a dedicated digital Toolkit.

3. The Cyrillic Palaeography Toolkit

The Cyrillic Palaeography Toolkit (CyPaT) is an online platform developed to support the description, structuring, processing, management, visualisation, and comparative analysis of codicological and palaeographical data derived from digitised images of Cyrillic manuscripts. Its primary aim is to provide a unified digital environment for the systematic and semi-automated analysis of script features, significantly enhancing research accuracy and efficiency.

After the user-initiated upload of the codex image, the analysis workflow begins with image preparation, including inclination correction and dimensional recalibration for precise proportional measurements. Integrated tools, such as the ‘sizer’, calculate the text field’s dimensions, while interlinear spacing is determined via a threshold line-by-line pixel analysis and segmentation process. Beyond recording the physical characteristics of the folio, which are incorporated into the ‘Manuscript’ object within the Repertorium, CyPaT provides extensive support for the analysis of script features, crucial for the ‘Palaeography’ section. Researchers can measure proportions between letter height and nib width, assess writing and text angles, and conduct detailed stroke analysis using a standardised palaeographic vocabulary.

To support this descriptive process, CyPaT provides a basic letter description template, which mirrors the general structure of individual letter strokes. In this way, the reconstruction of the precise form of each letter is made possible by proportionally calculating key features such as letter height, stroke width, and the various angles of pen inclination. The templated vector definition of the letters using Bezier curves produces a precise set of structured metadata which can be used for future analysis and for machine learning purposes.

A recently integrated and currently developing functionality is the ‘Match Pattern’ feature, which enables the selection of a single letter to be tracked across a manuscript folio, identifying its recurring instances and assessing their degree of similarity. This tool not only allows researchers to monitor the consistency of script within a manuscript, but also proves especially effective in identifying shifts in stylistic conventions, graphic variants, and changes in scribal practices. Such analysis can yield critical insights into the dynamics of collaborative work, revealing moments where multiple scribes may have contributed to the text’s composition, thereby offering a deeper understanding of the manuscript’s production process.

The functionalities described above are made possible through a carefully designed technological framework, built with a focus on accessibility, interoperability, and performance. The main technological strategy followed is to stick to open-source technologies in order to publish our software as a free and open-source product. Thus, more users may have access to the software, use its features and give feedback. The CyPaT was designed and implemented as a web-based application, allowing use from any modern web browser without the need for additional installation on the users’ personal computers.

The integration of CyPaT with the Repertorium constitutes a key future step toward a unified, data-rich environment for the digital analysis of South Slavic manuscript heritage. As CyPaT

data are currently entered manually into the Repertorium, full interoperability remains a central development goal. Ongoing work focuses on automating essential functions and building an annotated letterform repository, thus enabling large-scale analysis, scribal attribution, and the reconstruction of transregional networks. More than a technical toolkit, CyPaT seeks to reshape the study of medieval Cyrillic culture in the digital age.

Acknowledgements

This research work was carried out and is funded partly by the Bulgarian Scientific Fund under the research project No. KP-06-N90/9/11.12.2024 “Development of the Cyrillic alphabet from the 9th to the 14th century in the South Slavic lands: research and digital presentation”. It is also supported by the Bulgarian Ministry of Education and Science, Scientific Programme “Enhancing the Research Capacity in Mathematical Sciences (PIKOM)”, No. DO1-67/05.05.2022.

References

- [Cyr25] Cyrillic Palaeography Toolkit (2025). <https://kopisti.kmnc.bg/toolkit/>
- [LGP-M*24] LUCHEV, D., GOYNOV, M., PANEVA-MARINOVA, D., PAVLOV, R., RANGOICHEV, K., ZLATKOV, L. “Digital Tools for Medieval South Slavic Manuscripts Research”. Proceedings of the International Conference on Computer Systems and Technologies 2024 (CompSysTech '24). 2024, 102-108. DOI: [10.1145/3674912.3674946](https://doi.org/10.1145/3674912.3674946)
- [P-MGZ*23] PANEVA-MARINOVA, D., GOYNOV, M., ZHELEV, Y., MONOVA-ZHELEVA, M., MITREVA, E., LUCHEV, D., and PAVLOV, R. “Full-fledged Access and Usability of Content in Digital Cultural Heritage Library: Approaches, Paradigms and Implementation”. ACM Journal on Computing and Cultural Heritage 17.1, Article 8 (March 2024), 12 pages. DOI: [10.1145/3631135](https://doi.org/10.1145/3631135)
- [P-MSR*22] PANEVA-MARINOVA D., SCARPA M., RIPARANTE, M., GOYNOV, M., LUCHEV, D., PAVLOVA, L., and ZLATKOV, L. “Machine-readable Descriptive Structures to Study Medieval South Slavic Scriptoria and Scribes”. TEM journal 11.2 (May 2022), 675–682. DOI: [10.18421/TEM112-22](https://doi.org/10.18421/TEM112-22)
- [RD23] RIPARANTE, M., and DAVIDOVIĆ, M. “The Case Study of the Serbian Copyist Ioan. On the Development of a Scientific Method for South Slavonic Palaeography”. Digital Presentation and Preservation of Cultural and Scientific Heritage, Conference Proceedings. Vol. 13. 2023, 127–136. DOI: [10.55630/dipp.2023.13.12](https://doi.org/10.55630/dipp.2023.13.12)
- [Rep25] Repertorium of South Slavonic Manuscripts and Copyists (X–XIV cc.) (2025). <https://kopisti.kmnc.bg/en>
- [Res25] Research Project “Fourteenth Century South Slavonic Scribes and Scriptoria (palaeographic attribution and online repertorium)” (2025). <https://kopisti14.kmnc.bg/>
- [Rip24] RIPARANTE, M. “A Proposal for the Palaeographic Analysis of Medieval South Slavonic Cyrillic Script”. Palaeobulgarica XLVIII.2 (2024), 113–127. DOI: [10.59076/2603-2899.2024.2.06](https://doi.org/10.59076/2603-2899.2024.2.06)