

# Machine-readable Descriptive Structures to Study Medieval South Slavic Scriptoria and Scribes

Desislava Paneva-Marinova<sup>1</sup>, Marco Scarpa<sup>2</sup>, Marta Riparante<sup>2</sup>, Maxim Goynov<sup>1</sup>,  
Detelin Luchev<sup>1</sup>, Lilia Pavlova<sup>3</sup>, Lubomir Zlatkov<sup>1</sup>

<sup>1</sup> *Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, Sofia, Bulgaria*

<sup>2</sup> *Cyrillo-Methodian Research Center at the Bulgarian Academy of Sciences, Sofia, Bulgaria*

<sup>3</sup> *Laboratory of Telematics at the Bulgarian Academy of Sciences, Sofia, Bulgaria*

**Abstract** – This paper outlines the development of machine-readable semantic structures, supporting the research and analysis of the manuscript works of the South Slavic scribes from the fourteenth century. Our hope is to reach and effectively exploit the vast amounts of knowledge about the digitized written heritage by providing means to semantically describe, structure, process, manage, visualize and analyze the data, greatly facilitating the research and study processes.

**Keywords** – Medieval South Slavic Manuscripts and Scribes, semantic description, data access and analysis, manuscripts similarity discovery.

## 1. Introduction

The fourteenth century was a period of extraordinary cultural development throughout the Balkan Peninsula, clearly expressed in the lands of the southern Slavs [1].

In particular, with the establishment of the Bulgarian and the Serbian kingdoms, as they called their states, the foundation was laid for a great cultural and political rise, clearly reflected in the progress and scale of their manuscripts production.

The research on the writing of our predecessors of that time is a long-standing goal of modern researchers of the Slavic languages. Their main objective is to collect data from primary medieval written sources and to present them as a reliable instrument for identifying the provenance and defining the manuscripts currently preserved in libraries and institutional archives. For the realization of the objective, it is essential to investigate the kind of information that could provide the knowledge, needed to attribute medieval texts to individual scribes (copyists, writers, calligraphers), as well as cast further light upon South Slavic scribes and calligraphic schools from the fourteenth century.

Collecting data from primary medieval written sources, determining the origin, defining and systematizing the manuscripts preserved in various repositories (viz. manuscript repositories of libraries located in Bulgaria, Serbia, Croatia, Montenegro, Northern Macedonia, Bosnia and Herzegovina, Russia, Ukraine, Romania and on Mt Athos, as well as in the monastery of Sinai and in different libraries in Western Europe) in terms of their palaeographic features are serious research tasks. These activities are greatly facilitated when modern IT instruments assist with efficient data access, analysis and acquisition of new knowledge [2]. The current research work presented in this paper aims to propose powerful IT tools for the study and research of handwriting, copyist and writing in general, providing means to describe, structure, process, manage, analyze and visualize manuscripts content. Each written resource is digitized and described according to a common descriptive model based on the main practices, methodical rules and

---

DOI: 10.18421/TEM112-22

<https://doi.org/10.18421/TEM112-22>

**Corresponding author:** Desislava Paneva-Marinova,  
*Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences, Sofia, Bulgaria.*


**Email:** [dessi@cc.bas.bg](mailto:dessi@cc.bas.bg)

*Received:* 17 March 2022.

*Revised:* 21 April 2022.

*Accepted:* 27 April 2022.

*Published:* 27 May 2022.

 © 2022 Desislava Paneva-Marinova et al; published by UIKTEN. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 License.

The article is published with Open Access at  
<https://www.temjournal.com/>

recommendations related to the catalog of manuscripts. Once entered into the digital repertoire, it is discoverable through various context-based search and intelligent curation services. Structural metadata are provided for detailed examination and exploration of each manuscript and its elements. There is also a similarity discovery between pairs of manuscripts based on metadata descriptions, by handwriting typology and palaeographic features (*viz.* letters and their elements writing out, patterns in writing out, *etc.*).

The goal is to reach and effectively exploit the vast amounts of knowledge about the digitized written heritage with providing tools for higher adaptability and effective interaction between the researcher and the digital repertoire.

## 2. Research Methodology

This study is based on the combination of research methods, approaches and techniques that are peculiar for the different types of scientific specialization of the participants, both in the field of humanities and in the computer sciences [3]. Respectively, it includes a new type of competence, based not only on knowledge, but also on specific communication skills, sharing of ideas, concepts and experience. The methods and techniques take into account the exceptional dynamics of the development of the field, the applied research in it and its interdisciplinarity. They track the main line in the research practices: data collection → synthesis&analysis → assumptions&inferences&new knowledge. The expected results are to be achieved as a result of several stages of analysis:

- Archaeographic, palaeographic and textual: data collection, *i.e.*, search and selection of manuscript documents;
- Comparative: extraction of scientific information, *i.e.*, description, systematization and expert assessment of the cultural-historical context and socio-political situation in Europe and in the Balkans in the fourteenth century;
- Reconstructive: generalization of the data, their electronic processing and creation of a common methodology for the identification of the South Slavic handwriting, copyist and writing during the Middle Ages;
- Informational: promotion in accordance with modern social and technological requirements and opportunities for the generated new knowledge.

The interdisciplinary research presupposes a more well-developed perspective and new results in the machine-readable descriptive structures to study Medieval south Slavic scriptoria and scribes.

## 3. The Experience in Manuscripts Cataloging and Description

The 1960s marked the first attempts towards compilation of catalogues of Slavic manuscripts in some countries, with the pioneering work [4] comprising mainly archaeographic data on Russian manuscripts. Research results on the description of manuscripts from former Yugoslavia [5], Bulgaria [6] and Mt Athos [7] followed, establishing the solid basis for the cataloging of specific local repositories. Of particular notice is the role of the International Information Centre for Balkan Studies, that employed a specialized “manuscript commission”, which defined the model-rules and terminology for manuscript cataloging. The data collection for a preliminary list of the Slavic manuscripts, created on the Balkans until the end of the fourteenth century was completed by prof. Vasya Velinova in 1989 is one of the first examples of a digital data base.

During the last decades, a lot of libraries around the world participate in projects for the digitalization and online cataloging of the Slavic manuscripts. Stand-alone catalogues for digital presentation of Slavic manuscripts have been created in the National Library of Russia, State Historical Museum in Moscow, Russia, the national libraries of Czechia, Poland, Slovenia and Bulgaria, as well as for several monasteries. It is also important to mention several projects, supporting the creation of digital catalogues of Slavic manuscripts, such as the project “National Academic Library and Information System” (*NALIS*, <http://digilib.nalis.bg/xmlui/>), “Slovo: Towards a Digital Library of South Slavic Manuscripts” (*Slovo*, <http://slovo-aso.cl.bas.bg/index.html>) [8], the project “SCRIPTA BULGARICA: Digital Library of Medieval Bulgarian Literature” (*SCRIPTA BULGARICA*, <http://scripta-bulgarica.eu/en/manuscript>) [9], *etc.*, as well as the scientific results achieved in the field by Tsypkin [10], Kotseva [11], Kapskii [12].

Nevertheless, at the present moment, there is not a single good *common method for semantic description of Slavic manuscripts, applicable for following scientific research*. In fact, most of the codices are presented only with a short description, accompanied by microform or digital photographs. Unfortunately, this way of presentation does not provide exhaustive information, but rather only bibliographic references to the description of the manuscripts. With regard to the Medieval South Slavic manuscripts, there are also superficial attempts in achieving scientific and applied results in the contemporary field *digital paleography* (also called computer assisted paleography, which aims at “the study of ancient handwriting supported and enhanced by digital technologies” [13]). For the sake of comparison, we

can mention the notable advances, achieved for medieval Latin manuscripts in projects such as DMMapp (<https://digitizedmedievalmanuscripts.org/>), Digital Scriptorium (<https://digital-scriptorium.org/>), e-Codices (<https://www.e-codices.unifr.ch/>), *etc.*, in terms of access to the data, browsing through the images of manuscripts, image processing, image retrieval, pattern recognition on a micro structure level, *etc.*

#### 4. Descriptive Model of South Slavic Medieval Written Heritage

In the context of the pursuits for the creation not only of a Slavic manuscript catalogue or database, but rather of a complete environment for storage, extraction and intelligent curation of manuscripts and metadata, the research project № KP-06-N50/4 30.11.2020 "Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)" [3] initiated the work on creation of a common descriptive model of South Slavic medieval written heritage, in order to enable machine processing of the metadata, describing a certain manuscript, its components, scribes, paleographic features, *etc.* In the framework of this study, several types of attributes for the identification and attribution of South Slavic medieval written heritage are identified.

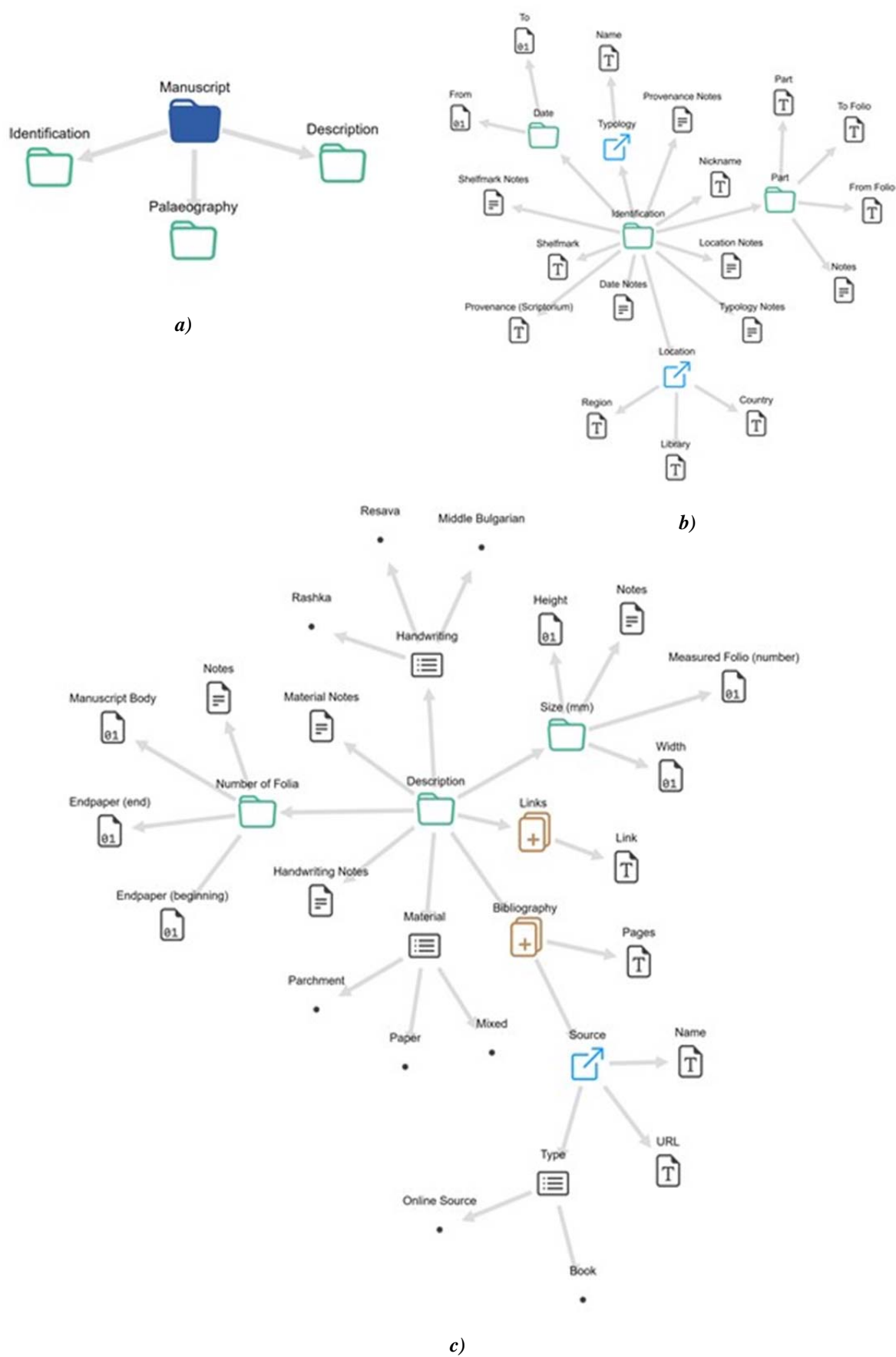
A key role plays the written resource, also referred to as a "Manuscript" object, which is digitized and described according to a descriptive model, underlying descriptors at the three levels of knowledge: "Identification", "Description" and "Paleography" (Figure 1.: Sub-figure 1-a). Sub-figure 1-b shows the "Identification" level containing common identifying aspects of a "Manuscript" object such as Nickname, Provenance (scriptorium), Location, Date, Shelfmark, *etc.* Sub-figure 1-c presents the "Description" level covering descriptors related to Handwriting, Material (Paper, Parchment, *etc.*), Size (mm), Number of Folia (Manuscript Body, Notes, Endpaper (beginning), Endpaper (end), *etc.*), Bibliography, *etc.* The level "Paleography" (see sub-figure 1-d) is the most important part for the identification of the paleographic features of the manuscripts. It covers descriptors such as Handwriting Size (mm), Handwriting Type, Handwriting Inclination (degree), Images, Copyists, Decoration, *etc.*

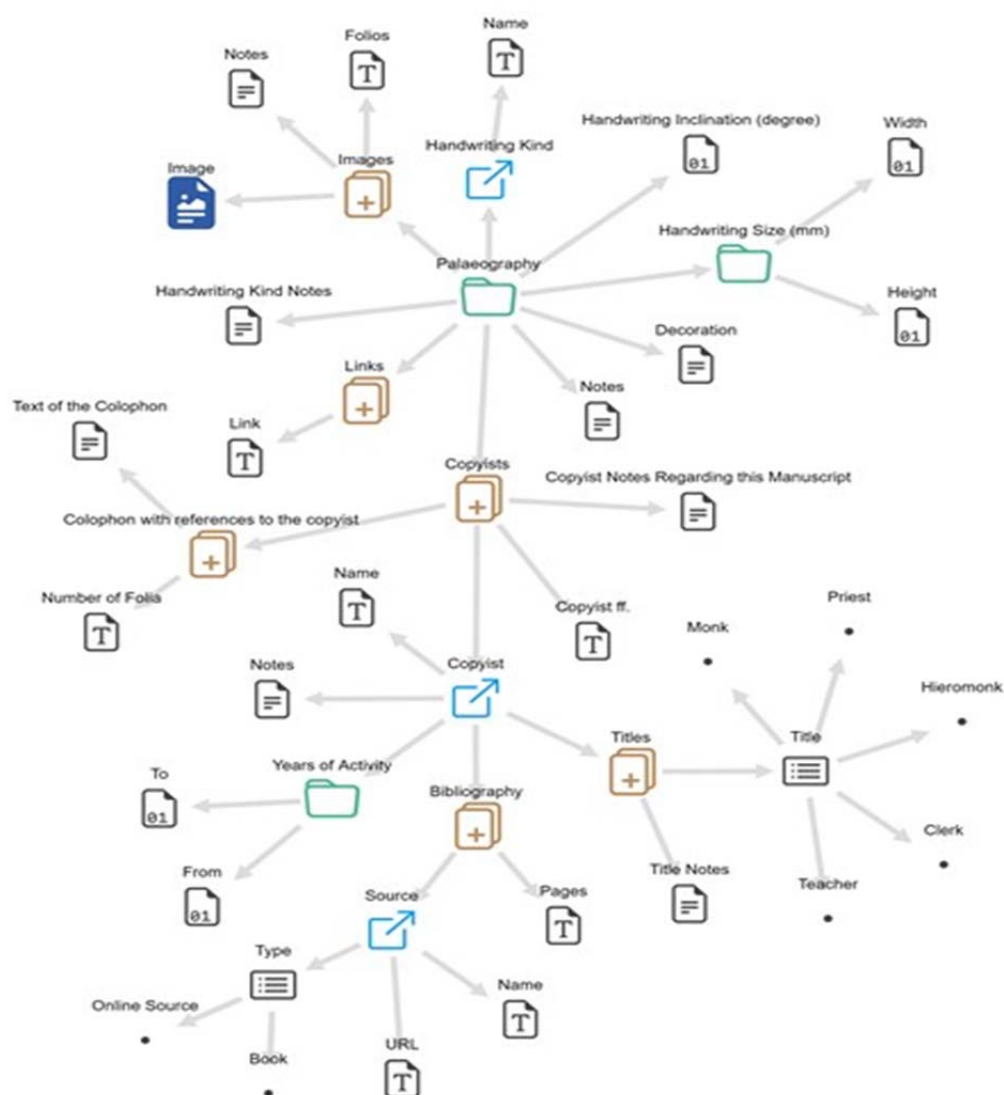
The descriptive model includes all types of relations typical for a conventional relational database, but implemented in a flexible and extensible way. The actual implementation uses nonrelational database and is developed in a way that allows the model to grow without the need of additional development. The one-to-many (1:N)

relation is implemented using the array element (A). Many-to-many relations (N:M) use a combination of arrays (A) and references (C). A reference element is just an address to an actual element (thus minimizing database redundancy). Sets of common elements (B) are used to keep things semantically well arranged. Data typed elements' purpose (D, E, F, G) is to guarantee the cleanness of data in the model.

The common descriptive model will allow:

- Description of Slavic medieval manuscript in various contexts – identification, content description and its paleographic features;
- Collection of data from primary medieval written sources, preserved in various repositories (*viz.* manuscript repositories of libraries located in Bulgaria, Serbia, Croatia, Montenegro, Northern Macedonia, Bosnia and Herzegovina, Russia, Ukraine, Romania and on Mt Athos, as well as in the monastery of Sinai and in different libraries in Western Europe);
- Effective exploitation of the vast amounts of knowledge about the digitized written heritage through the creation of a digital repertoire for these manuscripts, available on the Web and providing collection, browsing, search, synthesis, analysis systematization, intelligent curation and data management;
- Exhibition of new (covert) knowledge for the manuscripts – determination of the origin, the school, the presumed scribes (copyists, writers, calligraphers), exploration of tendencies, changes, influences, specification of currently unidentified paleographic features, organization of manuscripts from diverse locations with a common author, correlation of manuscript(s) and presumed/common author, school, *etc.*, identification of handwriting style, used decoration elements, *etc.*;
- Exhibition of new (covert) knowledge for the scribes (copyists, writers, calligraphers) – relation/association with manuscripts with unidentified authors (currently, the number of Slavic manuscripts from the fourteenth century with a precisely identified author is negligible, which poses a serious challenge in the field and is task of serious research importance), specification of handwriting styles, *etc.*, association of authors with known schools, introduction of new calligraphic schools, *etc.*;
- Implementation of a strong fundament for the provision of structured data to an upcoming machine learning model. Using the power of artificial intelligence, such a model could support researches in their analysis.





*d)*

*Agenda:*








	A	Array of objects (descriptors)
	B	A set of common descriptors
	C	Reference to model
	D	Single line text descriptor
	E	Multiline text descriptor
	F	Numeric value
	G	List of values

Figure 1. Descriptive model of the object "Manuscript": a) Main descriptive levels b) "Identification" level; c) "Description" level; d) "Paleography" level

**Palaeography**

**Handwriting Kind Uncial**

Handwriting Kind Notes

Български English

**Copyists (1)**

**Copyist Iakov Pant. 9, Pant. 84, Hil. 80, Hil. 422, Hil. 471, Hil. 612**

Copyist Notes Regarding this Manuscript

Български English

Copyist ff.

Copyist ff.

Colophon with references to the copyist (0)

**Handwriting Size (mm)**

Height

Height

Status: Draft

Save Delete More

**Related Manuscripts**

Greece, Holy Mount Athos, Hilandar Monastery, 612

Greece, Holy Mount Athos, Hilandar Monastery, 612

Greece, Holy Mount Athos, St Panteleimon Monastery, 84

**Relation Notes:** Turilov 2016 reports that the only folio of manuscript Pant. 84 belonged to manuscript Hil. 612

Figure 2. Part of the form for entering of metadata for the object "Manuscript"

References for Богдановић, Д., Каталог ћирилских рукописа манастира Хиландара, Београд 1978

Title	Type
<input checked="" type="checkbox"/> Lavrentiy Hil. 259, RNB F.I.376, Pog. 1054, Zogr. 2, Zo4gr.. 109, Zogr. 83	Copyist
<input checked="" type="checkbox"/> Greece, Holy Mount Athos, Hilandar Monastery, 47	Manuscript
<input checked="" type="checkbox"/> Greece, Holy Mount Athos, Hilandar Monastery, 1	Manuscript
<input checked="" type="checkbox"/> Greece, Holy Mount Athos, Hilandar Monastery, 9	Manuscript
<input checked="" type="checkbox"/> Greece, Holy Mount Athos, Hilandar Monastery, 18	Manuscript
<input checked="" type="checkbox"/> Roman 1 Hil. 3/1, Hil. 9, Hil. 160, Wuk. 49, Pant. 79, GIM-Sin.82	Copyist
<input checked="" type="checkbox"/> Martinian Hil. 47, Hil. 140/142, Hil. A6/22	Copyist
<input checked="" type="checkbox"/> Георги Радослав Hil. 1, Pant. 2, Zogr. 51, Grig. 11, IV, Grig. 12, II, NBIV 140 (602)	Copyist
<input checked="" type="checkbox"/> Anonymous Hil. 18	Copyist
<input checked="" type="checkbox"/> Greece, Holy Mount Athos, Hilandar Monastery, 15	Manuscript
<input checked="" type="checkbox"/> Greece, Holy Mount Athos, Hilandar Monastery, 13	Manuscript

Figure 3. Manuscripts and copyists, studied by D. Bogdanovich in "Cyrillic Manuscripts of Hilandar Monastery" [15]

## 5. Effective Access and Management of Digitized Written Heritage Assets

Under the research project “Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)” we have also started building an environment (digital repertoire) to explore collections of medieval south Slavic manuscripts.

Each object included in the digital repertoire (at this stage mainly digitized written resources, *viz.* manuscripts, copyists) is accessed and intelligently curated (browsed, discovered, previewed, examined for similarities, related to other objects, *etc.*) with automated services. The *repertoire* is a web-based software environment that provides the following main functional components: *metadata management and presentation* functional module, *advanced services* and *administrative services* that are linked to a *media repository* and a *user data repository*. It follows the team's know-how acquired in creating digital cultural content management systems [14] such as digital libraries, virtual museums, galleries, *etc.*

The functional module for metadata presentation and management reflects the primary tasks related to the creation and administration of metadata for objects (digitized written resources, copyists, *etc.*): adding (also object/image annotation), storing, editing, and deleting metadata; searching, selecting (filtering), accessing, viewing/displaying metadata. When creating the metadata, functionalities are provided to optimize the input, including tree structure of the annotation template, reuse of metadata, suggestion of already entered metadata values, auto completion, *etc.* Figure 2. presents the form for entering a “Manuscript” metadata, based on the already discussed common descriptive model of South Slavic medieval written heritage. Manuscripts, related to the current object, are shown on the right.

The search for objects is performed based on their common (one or more) descriptive characteristics, depending on the specific needs and research tasks. A module for the grouping of objects into collections (thematic, time and space dependent, objects with common paleographic features, *etc.*) is modelled and set to be implemented in the near future. Additionally, a module for creation, management and use of a dictionary of specific terms included in the description (metadata) of the stored objects, is available. The searching and tagging of a term (word form) is implemented automatically throughout the database records.

A positive side effect of the model could be observed during the data input, *e.g.*, *sources* (book titles or URLs), describing manuscripts or copyists. In this scenario, sufficient data input provides the opportunity for creating a list of contents of the *objects* studied in a certain source. Figure 3. shows a list of manuscripts and copyists, studied by D.

Bogdanovich and described in “Catalog of Cyrillic Manuscripts of Hilandar Monastery” [15].

Such connectivity allows updates to the inventory of the available manuscripts by a certain copyist, all manuscripts in a given location, all manuscripts with matching palaeographic features, *etc.* The semantic association of the descriptors enables a graph-like (as opposed to a more simplistic tree-like) structuring of the model.

Currently, a primary prototype of the digital repertoire has been created. The metadata input (in Bulgarian and English) account for more than four hundred South Slavic fourteenth century manuscripts.

## 6. Conclusion

The fourteenth century was a period of high importance for the cultural and historical development of the Balkans, which ultimately influenced the cultural dynamics of Europe in its entirety.

Written accounts on the endeavors of the South Slavic scribes and their monastic scriptoria in a period of vast cultural advancements in the region provide an indispensable information resource on an important part of the development of fourteenth century Europe. Their compilation, systematization, digitalization and curation could assist the researchers in assessing the knowledge gained from medieval manuscripts in new light, contributing to insights about the scientific and cultural policies of the states in the region, and Bulgaria in particular.

We propose a fully accessible online digital repertoire storing information on the works, the handwriting style and the biographies of scholars from the period, whose influential contributions to the developments of the South Slavic language, script and culture. The implementation of a strong foundation of structured data could lead us to both better theoretical understanding of the research question emerging in the digitized collections field, but also to much more practical contributions, *e.g.*, potential machine learning models, supporting the researchers in their analytical work.

Finally, we hope that our work could contribute to the preservation and dissemination of the knowledge of an important part of the cultural and historical heritage of Southeastern Europe. The efforts to revitalize these important information resources, often at risk of damage and loss, is also part of our continuous work towards preserving the Bulgarian, and in general the South Slavic, history, culture and identity – among others, a priority area in the Bulgarian National Strategy for Development of Scientific Research 2017-2030 and in accordance with the thematic fields set up in the European Framework Programme for Research and Innovation Horizon 2020, the basis of the new Horizon Europe (2021-2027).

## Acknowledgements

This research is supported by the Bulgarian Scientific Fund under the research project № KP-06-N50/4 30.11.2020 "Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)" (website: <https://kopisti14.kmnc.bg/en/>).

## References

- [1]. Scarpa, M., Blanchet, M.-E., & Batalova, S. (2021). *Religious bloom. Bulgaria 13th – 15th c.* Sofia: Cyrillo-Methodian Research Centre at the Bulgarian Academy of Sciences.
- [2]. Musakova, E. (2015). *Letter and writers in Bulgarian manuscripts until the end of the 17th century.* Sofia: National Academy of Arts.
- [3]. Scarpa, M., Burlacu, C., & Paneva-Marinova, D. (2021). New digital life for fourteenth century South Slavonic scriptoria. In *Digital Presentation and Preservation of Cultural and Scientific Heritage, Conference Proceedings.* (Vol. 11, pp. 57-64). Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences.
- [4]. Shelamanova, N. B. (1966). Preliminary list of Slavic-Russian manuscripts of the 11th-14th centuries stored in the USSR. In *Archaeographic Yearbook 1965* (pp. 177-272).
- [5]. Bogdanovich, D. (1982). *Inventar ćirilskih rukopisa u Jugoslaviji (XI–XVII veka).* Beograd: Serbian Academy of Sciences and Arts.
- [6]. Hristova, B., Karadzova, D., & Ikonomova, A. (1982). *Bulgarian manuscripts from 11th to 18th century preserved in Bulgaria. Consolidated Catalog, Vol. 1.* Sofia: St. St. Cyril and Methodius National Library.
- [7]. Turilov, A. A., & Moshkova, L. V. (2016). *Catalog of Slavic manuscripts of Athos monasteries / Catalog of Slovenian manuscripts of Svetogorsk manastir.* Beograd: Chigoja Shtampa.
- [8]. Miltenova, A., & Tomova, E. (2008). Slovo: Towards a digital library of South Slavic manuscripts research cooperation and networking between Austria and South-Eastern Europe. *Scripta & e-Scripta*, 6, 9-22.
- [9]. Miltenova, A. (2018). Scripta Bulgarica: Digital Library of Medieval Bulgarian Literature. *Slavistica Vilnensis*, 63, 309-326.
- [10]. Tsypkin, D. (2015). Poniatie pocherka v izuchenii russkogo istoricheskogo pisma. K probleme pazrabotki metodologii pocherkovedcheskogo analiza drevnerusskih rukopisei. *Trudy Otdela drevnerusskoj literatury*, LXVI, 836-866.
- [11]. Kotseva, E. (1985). Development of the Bulgarian Cyrillic alphabet in the XIV-XV centuries. In *Stanchev, K. Velcheva B. (Eds.). Slavic paleography and diplomacy, 2. Lectures from the second seminar on Slavic paleography and diplomacy and reports from the scientific conference "The decoration of the Slavic manuscript to the XVIII century"* (pp. 23–47). CIBAL.
- [12]. Karskii, E. F. (1979). *Slavic Cyrillic paleography.* Moskva: Nauka.
- [13]. Ciula, A. (2017). Digital palaeography: What is digital about it?. *Digital Scholarship in the Humanities*, 32(suppl\_2), ii89-ii105.
- [14]. Paneva-Marinova, D., Goynov, M., & Luchev, D. (2017). *Multimedia Digital Library: Constructive Block in Ecosystems for Digital Cultural Assets: Basic Functionality and Services.* Lap Lambert Academic Publishing.
- [15]. Bogdanović, D. (1978). *Katalog ćirilskih rukopisa manastira Hilandara* (Vol. 1). Srpska akademija nauka i umetnosti, Narodna biblioteka Srbije.