

Third Information Day: Research Infrastructure Services in the Humanities and Social Sciences

Desislava Paneva-Marinova^[0000-0001-5998-687X], Radoslav Pavlov,
Detelin Luchev^[0000-0003-0926-5796], Maxim Goynov,
Lubomir Zlatkov, Nikolay Noev^[0000-0003-3290-1439]

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

dessi@cc.bas.bg, radko@cc.bas.bg,
dml@math.bas.bg, goynov@gmail.com,
lyubcho@gmail.com, nickey@math.bas.bg

Abstract. The main results of the work of the team of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences on the research projects CLaDA-BG – the Bulgarian National Interdisciplinary Research e-Infrastructure for Resources and Technologies in Favor of the Bulgarian Language and Cultural Heritage, Part of the EU Infrastructures CLARIN and DARIAH and No KP-06-N50/4 30.11.2020 “Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)”, will be presented during the Information Day within the DiPP2024 conference. The main emphasis will be placed on developments and implementations of CultIS – a web-based software platform for intelligent digital management and presentation of large data sets and knowledge from the fields of culture, humanities, and social sciences.

Keywords: CLaDA-BG, CultIS, Research Infrastructure, Digital Humanities, Digital Content Management Systems, Intelligent Content Curation.

1 Introduction

Since 2000, the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences (IMI-BAS) has had enduring research interests in digital libraries and their applications, *incl.* innovative methods and tools for acquiring and presenting knowledge of various objects of interest (cultural, natural, *etc.*); tools for access and improved use of the digital cultural heritage, *etc.*

The present IMI-BAS development is the web-based platform for intelligent cultural content management *CultIS* (<https://cultis.math.bas.bg/en>) representing a complex web-based humanities and social sciences data storage, retrieval and curation environment containing a rich set of technologies aiming to support a great variety of digital cultural units. Its powerful features and components provide flexible storing, interactive virtual presenting of objects and collections, standard and complex search and grouping

in various intersections, flexible management and structuring of metadata, data, and objects, multilayer indexing of data using dictionaries, *etc.*, thus making it suitable for a number of diverse applications. The basic prototype of *CultIS* is able to store and manage different types of digitized copies of cultural heritage objects, including text, graphics, video, audio, 3D formats, or other media objects, as well as the relevant metadata. *CultIS* provides a full metadata managing stack for creating, storing, editing, deleting metadata, backing up, while version control is covered by the functional module for managing and presenting metadata. The module is also responsible for user searches, data filtering, access management, and metadata presentation over a variety of user devices.

The *CultIS* core platform is implemented for the needs of the Central library of the Bulgarian Academy of Sciences, the National library “Ivan Vazov” – Plovdiv, the National Statistical Institute, the Regional library “Peyo Yavorov” – Burgas, the Sofia University Library (in progress), *etc.* It is an infrastructure component of the CLaDA-BG, the Bulgarian National Interdisciplinary Research e-Infrastructure for Resources and Technologies in favour of the Bulgarian Language and Cultural Heritage, part of the EU infrastructures CLARIN and DARIAH (2018-2027), whose mission is to create a national technological infrastructure for resources and technologies for linguistic and cultural heritage.

Another key research project of the IMI-BAS team is the “Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)” (No KP-06-N50/4/30.11.2020) supported by the Bulgarian Scientific Fund. From a paleographic perspective, it aims to collect data from medieval primary written sources and providing that data to the scientific community as a reliable tool for origin resolution, definition, and systematization of the manuscripts, preserved currently in the book repositories (Scarpa et al., 2021). The project is concentrated on the exploration of new sources and their consolidation with existing sources, thus contributing to the methods for reliable identification of specific medieval scribes, based on the information on South Slavic copyists from the fourteenth century – a particularly significant period for the cultural and historical development on the Balkans, influential on the cultural processes in Europe as a whole.

During the Third Information Day: Research Infrastructure Services in the Humanities and Social Sciences current and outgoing research results of the mentioned projects will be presented.

2 IMI-BAS’ Reseach Results in CLaDA-BG e-Infrastructure

CLaDA-BG (<https://clada-bg.eu/>) is a leading initiative in Bulgaria, aiming to promote the innovative use of resources in order to encourage the sustainable development of European cultural landscapes in a digital environment (Simov, 2019). The mission of CLaDA-BG is to establish a national technological infrastructure of language, cultural and historic heritage (CHH) resources and technologies which to provide public access to language and CHH resources, tools for Bulgarian language processing and tools for

access and management of CHH datasets for various societal tasks, targeted at wide audience. The infrastructure aims to support especially researchers in Arts, Humanities and Social Sciences to process Bulgarian language texts and CHH datasets necessary for their research. As a strategic task, CLaDA-BG seeks to preserve the Bulgarian language, cultural and historical heritage and to provide wide access to language resources, CHH datasets and technology to all potential users – the research community, education, and industry.

Over the past year, the IMI-BAS team has continuously worked on the development and implementation of the *CultIS* environment for humanities and social sciences data storage, retrieval, and curation in the framework of CLaDA-BG, consistent with our goal of providing flexible and efficient access to multimedia representations of cultural and historical artifacts, supporting a variety of forms and formats of digital information content and rich interaction functionality (Luchev et al., 2021).

CultIS is a web-based software platform for intelligent digital management and presentation of large datasets and knowledge in the fields of culture, humanities, and social sciences (<https://cultis.math.bas.bg/>). The main functionalities of the system are aimed at interactive virtual presentation of objects and collections, standard and complex search and grouping in various intersections, flexible management and structuring of metadata, data, object, multilayer indexing of data using dictionaries, *etc.*

The development of *CultIS* is aimed at achieving a high degree of adaptability in terms of building specific designs and user interfaces for the needs of specific environments. It has possibilities for implementation of additional modules and functionalities according to application areas and for importing data from various systems and formats.

The current implementations of *CultIS* have been created with the support of the National library “Ivan Vazov” – Plovdiv (Paneva-Marinova, et al., 2022), the Regional library “Peyo Yavorov” – Burgas (Paneva-Marinova, et al., 2024), the Central Library of the Bulgarian Academy of Sciences (Goynov, et al., 2023), *etc.* There is huge interest in *CultIS* from important institutions such as the National Statistical Institute, the Sofia University Library, and regional libraries and museums, some of which have started joint implementation activities.

A key *CultIS* implementation is the digital library of the National Library “Ivan Vazov” – Plovdiv. It stores and manages the digital analogues of cultural heritage objects presented in text (via .pdf files, fully corresponding to book media), graphic, video, audio formats, or other media objects, as well as the relevant metadata. Specific services for intelligent content curation and context-based digital library content usage for research and e-learning, as a part of this module, include functionalities closely related to the content provider needs, such as advanced collection creation, management and curation (thematic collections, time dependent collections in a calendar structure, *etc.*), search in the text media objects, advanced objects preview and ordering, different device support, *etc.* The National Library “Ivan Vazov” – Plovdiv prototype provides specific services for indexing Portable Document Format (PDF) objects, providing the opportunities for full-text search in the objects’ content, *etc.* The Digital Library – National Library Ivan Vazov – Plovdiv is available at: <https://digital.libplovdiv.com/>.

The implementation of *CultIS* for the needs of the Central Library of BAS began with the work on one of the noteworthy collections of CL-BAS, associated with the

international congresses of Bulgarian studies from the 1980s (the *Bulgarica* Collection) and consisting of more than nine hundred scientific papers and reports. Because of the challenges during the implementation of *CultIS* for the needs of the Central Library of BAS, related to the migration of data from an older platform used as a digital storage and from separate archive units, new modules were designed and implemented.

The digital collections of the papers of the First Second and Third International Congresses of Bulgarian Studies are available at: <https://clbas.bg73.net/>.

3 Digital Tools for Medieval South Slavic Manuscripts Research

In order to assist the research and analysis of the manuscript works of the South Slavic scholars of the fourteenth century, IMI-BAS with the cooperation with the Cyrillo-Methodian Research Centre at the Bulgarian Academy of Sciences works on an *online accessible digital repertorium* (<https://kopisti.kmnc.bg/bg>), developed to store and preserve information about valuable manuscripts and copyists (scribes) with significant contribution to the development of the South Slavic languages, writings and culture. The goal is the attainment and efficient utilization of the vast amounts of knowledge on the digitalized written heritage, and the provision of tools for higher adaptivity and efficient interactions between the researcher and the digital repertorium. The information about the Slavic manuscripts and copyists from the fourteenth century is processed, analyzed and published in the online repertorium, based on an established theoretical framework and a working scientific methodological approach. From a research perspective, the collection and classification of the required data for reliable identification of genuine medieval calligraphers aims at the derivation of new knowledge about the organization and activities of the calligraphic schools and local workshops, the primary centers for calligraphic education and production. Each literary resource is digitized and described according to the descriptive model based on the available metadata. Once added to the digital repertorium, it could be found through a variety of context-based services for browsing and intelligent curation. The tools provided offer the means for detailed research and analysis of each glyph and its elements, detection of similarities between manuscripts (*e.g.* inscription of letters and their elements), support for collaborative, simultaneous work by researchers in the system, and management of the process of description and structurization of data.

The project includes machine-readable and -processable semantic structures development (*i.e.*, *Descriptive model of the Medieval South Slavic literary heritage from the fourteenth century*), allowing efficient digital access, semantic description, structuring, automated processing, management, visualization, and analysis of immense volumes of scientific knowledge, data, and metadata of the invaluable literary heritage, accessible and presented in the digital repertorium. The model is mainly focused on the specific manuscript and its components, copyists, writing and its paleographic features.

4 Conclusions

The development and implementation of tools for intelligent digital management and presentation of large datasets and knowledge in the fields of culture, humanities and social sciences is very important for current researchers in the humanities and social sciences. *CultIS* offers specific and individual solutions for the needs of various libraries and museums in terms of intelligent digital management and presentation of national heritage and knowledge therein. The results of the research activity are aimed at the partners of the CLaDA-BG network, but they can be used and applied also in other cultural and scientific environments.

Discovering new directions for the development of established digital tools is essential to achieving new research outcomes in the humanities.

Acknowledgements.

This research work was carried out and is funded partly by CLaDA-BG, the Bulgarian National Interdisciplinary Research e-Infrastructure for Resources and Technologies in favour of the Bulgarian Language and Cultural Heritage, part of the EU infrastructures CLARIN and DARIAH, Grant numbers DO1-167/28.07.2022 and DO1-301/17.12.2021, <https://clada-bg.eu/bg/> and the Bulgarian Scientific Fund under the research project No KP-06-N50/4/30.11.2020 “Fourteenth Century South Slavonic Scribes and Scriptoria (Palaeographical Attribution and Online Repertorium)” (web page: <https://kopisti14.kmnc.bg/>).

References

- Goynov, M., Luchev, D., Paneva-Marinova, D., Najdenova, S., Zlatkov, L., Pavlova, L., & Pilege, E. (2023). Digital Revival of the Bulgarica Collection of the Central Library of the Bulgarian Academy of Sciences. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 13, 77–86. <https://doi.org/10.55630/dipp.2023.13.7>
- Luchev, D., Goynov, M., Paneva-Marinova, D., Stoykov, J., & Pavlova, L. (2021). Synergy of National Cultural Heritage and Technology. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 11, 281–286. <https://doi.org/10.55630/dipp.2021.11.26>
- Scarpa, M., Burlacu, C., & Paneva-Marinova, D. (2021). New Digital Life for Fourteenth Century South Slavonic Scriptoria. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 11, 57–64. <https://doi.org/10.55630/dipp.2021.11.5>
- Simov, K. (2019). Integrated Language and Knowledge Resources for a Bulgarian-Centric Knowledge Graph. *Digital Presentation and Preservation of Cultural and Scientific Heritage*, 9, 65–74. <https://doi.org/10.55630/dipp.2019.9.5>

- Paneva-Marinova, D., Goynov, M., Zhelev, Y., Monova-Zheleva, M., Mitreva, E., Luchev, D., Pavlov, R., & Pavlova, L. (2024). Full-Fledged Access and Usability of Content in a Digital Cultural Heritage Library: Approaches, Paradigms, and Implementation. *Journal on Computing and Cultural Heritage*, 17(1), Article 8 (March 2024), 1-12. <https://doi.org/10.1145/3631135>
- Paneva-Marinova, D., Minev, D., Krachanov, I., Luchev, D., Goynov, M., Zlatkov, L., Pavlov, R., & Pavlova, L. (2022). The Power of Intelligent Content Curation and Context-Based Digital Library Content Usage for Research and E-Learning (The Case of National Library “Ivan Vazov” – Plovdiv, Bulgaria). In *INTED2022 Proceedings. 16th International Technology, Education and Development Conference* (pp. 4013-4021). <https://doi.org/10.21125/inted.2022.1097>

Received: March 25, 2024

Reviewed: April 15, 2024

Finally Accepted: April 30, 2024