

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

Desislava Paneva-Marinova¹[0000-0001-5998-687X], Maxim Goynov¹,
Detelin Luchev¹[0000-0003-0926-5796], Yanislav Zhelev^{1, 2}[0000-0003-2783-5617],
Mariya Monova-Zheleva^{1, 2}[0000-0001-8910-2502], Radoslav Pavlov¹,
Lubomir Zlatkov¹, Nikolay Noev¹[0000-0003-3290-1439], Lilia Pavlova³

¹ Institute of Mathematics and Informatics, Bulgarian Academy of Sciences, Sofia, Bulgaria

² Burgas Free University, Burgas, Bulgaria

³ Laboratory of Telematics, Bulgarian Academy of Sciences, Sofia, Bulgaria

dessi@cc.bas.bg, qoynov@gmail.com, dml@math.bas.bg,
yanislav@zhelev.com, mariya@zhelev.com, radko@cc.bas.bg,
lyubcho@gmail.com, nickey@math.bas.bg, pavlova.lilia@gmail.com

Abstract. During the information day, the results of the work of the IMI-BAS team on the project 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, will be presented: the development of the Humanities and Social Sciences Data Storage, Retrieval and Curation Environment (CHCS-DSRCE) and its implementation in the Digital Library "Virtual Encyclopedia of Bulgarian Iconography" and for the needs of the "Ivan Vazov" Regional Library in Plovdiv and "Peyo Yavorov" Regional Library - Burgas.

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

1 Introduction

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, 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.

The team of the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences (IMI-BAS), as a technological partner, works on the creating of technologies and tools to unify the processes of access, preservation and use of Bulgarian language and cultural historical content in accordance with the established best practices and regulations in the field. The development of The *Humanities and Social Sciences Data Storage, Retrieval and Curation Environment (CHCS-DSRCE)* aims to provide 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 functionality for interacting with it (Luchev, Goynov, Stoykov, & Pavlova, 2021).

2 The Humanities and Social Sciences Data Storage, Retrieval and Curation Environment

During the third stage of the project, the team of IMI-BAS is working on building the technological base and necessary infrastructure components of the Research E-infrastructure CLaDA-BG. Research and activities focus on the development of solutions for intelligent digital management and presentation of objects and knowledge of the national cultural heritage.

Work is being done on the development of the architectural solution of the environment for storing, retrieving and curating data from the field of humanities and social sciences, following its previous developments and taking into account the peculiarities of its implementation in various areas of the Bulgarian cultural and historical heritage. Emphasis is placed on storing, retrieving, and curating data and metadata about target objects. The development is aimed at creating means and technologies to help and serve the needs of cultural and scientific institutions in the field of humanities and social sciences, users of the resources of CLaDA-BG, libraries, museum, etc.

The CHCS-DSRCE is a web-based software environment that provides the following functional components: a *metadata management* and *presentation functional module*, a *metadata model management* module, *administrative services* that are linked to a *media repository* and a *user data repository*.

After the completion of the basic prototype of a web-based software environment CHCS-DSRCE, the stage of experimental and test implementations was launched, carried out independently using digital cultural resources of IMI-BAS and together with partners in the project - providers of information content.

3 The Digital Library "Virtual Encyclopedia of Bulgarian Iconography"

The IMI-BAS team continued the work on analyzing specific sources from the field of Bulgarian iconographic art with the aim of extracting knowledge through the knowledge network. The work is being done on the experimental and test implementation of the environment for storing, retrieving and curating data from the field of Bulgarian iconographic art. Software development includes all the basic modules of the base environment, extended with *specific functionalities and structures*:

- functionality allowing a user with an editor or administrator role to create specific search forms in the various types of objects (without the need for specific technical knowledge; in the base version, such shapes are only created programmatically);

- functionality allowing an administrator to define a specific presentation of individual types of object categories, as well as of the objects themselves;
- functionality for separating layers of full-text search - currently implemented layers - metadata layer and text content layer;
- functionality for simultaneous correction of metadata in a large number of objects (batch processing);
- functionality of simultaneous reindexing, validation, generation of a representative image (thumbnail), data consistency check of multiple objects, at the request of an administrator or as part of a data migration process, etc.

The developed in co-operation between IMI-BAS and Institute for Computer Science and Control (SZTAKI), Hungary, virtual walk in the Church of Nativity in Arbanasi is integrated with the Digital Library "Virtual Encyclopedia of Bulgarian Iconography" with the help of hotspots, which are placed at the selected icons of the virtual walk, and by clicking on a hotspot, the image and description of the icon in Digital Library "Virtual Encyclopedia of Bulgarian Iconography" is presented online in Bulgarian and English (Zsolt, et al., 2022).

Work is underway to prepare the descriptions of selected iconographic objects and iconographic content from the Burgas region, which will be included this year in the knowledge base of the knowledge network implemented through the INCEPTION system.

The Digital Library "Virtual Encyclopedia of Bulgarian Iconography" is available at: <https://bidl.math.bas.bg/en>.

4 The Environment for Storing, Retrieving and Curating Data from the Field of Humanities and Social Sciences for the Needs of the "Ivan Vazov" Regional Library in Plovdiv

The IMI-BAS team is working on the experimental and test implementation of the environment for storing, retrieving and curating data from the field of humanities and social sciences for the needs of the "Ivan Vazov" Regional Library in Plovdiv (Paneva-Marinova, et al., 2022). "Ivan Vazov" Regional Library in Plovdiv owns a rich, multi-functional fund of over 1,460,000 library items - scientific literature and fiction; manuscripts, rare and valuable publications; rich reference fund; Bulgarian and foreign periodicals; audio-visual documents; original works of art, *etc.*, and digitization activities have not stopped since 1996 until today. The large volume of library items requires well-defined functionality for searching, ordering, flexible access, and visualization, meeting the needs of librarians and viewers. The experimental and test implementation of the environment for storing, retrieving and curating data from the field of humanities and social sciences for the needs of the "Ivan Vazov" Regional Library in Plovdiv includes all the basic modules of the base environment, expanded with *specific functionalities and structures*:

- construction of formal schemes for the description of various collections of cultural and historical objects, incl. for periodicals, books, photographs,

geographical maps, audio-visual objects, etc.; formalization and creation of semantic links between objects;

- full-text search services in the meta-descriptions and in the actual digital objects; PDF Document Processing – building functionalities for indexing the content of PDF documents for the purpose of effective full-text search;
- intelligent content curation (structuring of periodicals in the form of a calendar;
- creation of services for visualization in a web-browser of digital correspondences of cultural-historical artifacts from the "periodicals" collection (pdf format);
- visualization of the collections and their artifacts - a specific (customized) presentation layer (user interface for arranging/displaying the objects in a calendar, ordered lists, etc.);
- functionality for discovering and tracking semantic connections between different digital cultural objects (data analytics) at the modelling stage;
- Audio/Video Formats Processing;
- protection of digital objects through complex watermarking technologies; limiting the indexing of search engines, (where necessary) etc.

A test version of the environment is available at: <https://nbiv.bg73.net/bg>.

5 Experimental and Test Implementation of the CHCS-DSRCE for the Needs of “Peyo Yavorov” Regional Library - Burgas

Work also began on the experimental and test implementation of the CHCS-DSRCE for storage, retrieval and curation of data from various areas of the Bulgarian cultural and historical heritage for the needs of "Peyo Yavorov" Regional Library - Burgas in connection with its association as a partner of CLaDA-BG. It is planned to gradually include various collections such as periodicals, printed materials (posters, cards), maps, manuscripts, etc. and to build specialized functionality for user needs. The collection of posters stored in the "Peyo Yavorov" Regional Library - Burgas is available at: <https://plakati.bg73.net/bg>.

6 Conclusions

The research activity of IMI-BAS within the framework of CLaDA-BG aims to consistently build a scientific and informational infrastructure integrating research, education, preservation, promotion and sustainable use of the national cultural heritage. It seeks to explore the potential of information technology to support the work of researchers in the humanities and social sciences by offering solutions for the intelligent digital management and presentation of national heritage sites and knowledge.

The next steps for the implementations of CHCS-DSRCE concern development of the environment for the needs of “Peyo Yavorov” Regional Library – Burgas. Before launching the live version of the environment for "Ivan Vazov" Regional Library in

Plovdiv, will be necessary to include in the front-end part the other types of objects – manuscripts, photographs, graphic publications, maps, audio visual content, *etc.* , and to perform security, load, stress and performance tests, to scale properly the hardware, to define proper backup policies, which will guarantee high availability of the environment and a good protection level for the users, content and its respective metadata.

Acknowledgements.

This research is partly supported by 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, Grant number DO01-301/17.12.2021.

References

- Luhev, 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*, Sofia, Bulgaria: Institute of Mathematics and Informatics - BAS, (pp. 281-286).
- Márkus, Z.L., Szkaliczki, T., Szántó, G., Veres, M. Weisz, Z., Luhev, D., Paneva-Marinova, D., Goynov, M., Pavlov, R., Pavlova, L. (2022). Virtual Walk in the Church of Nativity in Arbanasi. *Digital Presentation and Preservation of Cultural and Scientific Heritage, 12*, Sofia, Bulgaria: Institute of Mathematics and Informatics - BAS, (pp. 63-73).
- Paneva-Marinova, D., Minev, D., Krachanov, I., Luhev, 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). *16th annual International Technology, Education and Development Conference, IATED*, (pp. 4013-4021). doi:10.21125/inted.2022.
- Simov, K. (2019). Integrated Language and Knowledge Resources for a Bulgarian-Centric Knowledge Graph. *Digital Presentation and Preservation of Cultural and Scientific Heritage. Vol. 9*, Sofia, Bulgaria: Institute of Mathematics and Informatics - BAS, (pp. 65-74).

Received: June 05, 2022

Reviewed: June 20, 2022

Finally Accepted: July 15, 2022

