

Synergy of National Cultural Heritage and Technology

Detelin Luchev¹[0000-0003-0926-5796], Maxim Goynov¹,
Desislava Paneva-Marinova¹[0000-0001-5998-687X], Jordan Stoykov¹, Lilia Pavlova²

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

²Laboratory of Telematics, Bulgarian Academy of Sciences, Sofia, Bulgaria
dml@math.bas.bg, goynov@gmail.com, dessi@cc.bas.bg,
jstoikov@shieldui.com, pavlova.lilia@gmail.com

Abstract. This paper presents the specification of the *Humanities and Social Sciences (CHCS) data storage, retrieval and curation environment* developed by the IMI-BAS team within the Bulgarian National Interdisciplinary Research E-infrastructure for Bulgarian Language and Cultural Heritage Resources and Technologies (CLaDA-BG). The architecture and functional components of the environment and the technologies used for the development of the environment are presented. The aim of the development is to provide and propose technologies and tools to unify the processes of access, preservation and use of cultural and historical content in accordance with established best practices and regulations in the field.

Keywords: Research e-infrastructure, Digital Humanities, Digital Content Management Systems.

1 Introduction

CLaDA-BG is the Bulgarian National Interdisciplinary Research E-Infrastructure for Bulgarian Language and Cultural Heritage Resources and Technologies. In the spirit of European CLARIN and DARIAH, the mission of CLaDA-BG is to establish a national technological infrastructure of language resources and technologies, and cultural and historical heritage resources and technologies (Simov, 2019). It is included in the National Roadmap for Science Infrastructure (2017-2023), adopted by Decision No. 354 of 29 June 2017 of the Council of Ministers. The CLADA-BG involves 15 organizations (scientific structures of BAS, scientific structures of universities, etc.), which according to their functions and responsibilities in the building of the infrastructure are divided into two groups 1) technology partners and 2) user partners (application partners). Technology partners, such as the Institute of Mathematics and Informatics at BAS (IMI-BAS), build and maintain the core technologies and resources necessary for the functioning of CLADA-BG. The application partners are the first users of the built services and resources of CLADA-BG and will demonstrate through sample applications the effectiveness of the built infrastructure and the direction of further development of the resources and technologies offered in CLADA-BG. CLADA-BG aims to

promote the innovative use of resources in order to encourage the sustainable development of European cultural landscapes in a digital environment.

A specific goal of CLADA is to lay the foundations for a network of cultural and historical heritage data and technologies that will ensure future free access for researchers and citizens to these technologies and their results, helping to bring them more widely into everyday life.

The research activities of IMI-BAS within the project aim to consistently build scientific and information infrastructure, integrating research, education, preservation, promotion and sustainable use of 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 cultural heritage objects and knowledge.

This paper presents the specification of the environment for data storage, retrieval and data curation in the field of the *Humanities and Social Sciences (CHCS)*, a core task of the IMI team at CLADA-BG.

2 Environment for Data Storage, Retrieval and Curation in field of Humanities and Social Sciences - Architecture and Features

The Humanities and Social Sciences (CHCS) Data Storage, Retrieval and Curation Environment architecture 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. Emphasis is placed on the storage, retrieval and curation of data and metadata for target objects. It is based on the experience and knowledge gained from previous developments of the IMI team on digital content management systems (including digital libraries, digital repositories, galleries, etc.) preserving the valuable Bulgarian cultural heritage: Bulgarian iconographic art, Bulgarian ethnographic and folklore artifacts, medieval and early modern Bulgarian texts about saints in combination with ethnological data and visual sources, etc.

The environment will store the digital analogues of cultural heritage objects presented in text (via .pdf files, fully corresponding to book media), graphic, video, or audio formats, as well as other media objects. The resources are digitized and made available by a project partner- an information content provider, compliant with the specific requirements for their creation.

The *Humanities and Social Sciences Data Storage, Retrieval and Curation Environment* 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* (Figure 1).

The module for management of the metadata model, includes a service for building the descriptive schemas (descriptor structure) for cultural objects, the so-called Model builder. The component manages the "object metadata data". It also manages relationships between complex and heterogeneous objects. The service is designed in a way that allows:

- extensibility of the created models,
- flexibility of models and their areas of application,
- reuse of a model in different areas,
- reflecting high model complexity,
- integration with other model building systems (import/export of models),
- Compatibility with the most common semantic web standards including Dublin Core and others,
- easy to use with UI/UX best practices,
- easy navigation and browsing of large volume models,
- device responsiveness built on Bootstrap.

The key component in the environment for storing, retrieving and curating data on objects in the field of humanities and social science is the *functional module for managing and presenting metadata*. It implements the basic activities related to the creation and management of metadata for cultural objects: adding, 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, suggesting already entered metadata values, auto completion, etc.

Based on the presented module will be realized the grouping of cultural objects into collections, the creation of collections (thematic, time and space dependent, etc.) and their attractive virtual exposure. The content provider partner in the project specifies the metadata categories that will be used for grouping the objects into collections in the library. Through the module is implemented the creation, management and use of a dictionary of specific terms included in the description (metadata) of the stored objects. The searching and tagging of a term (word form) is implemented automatically throughout the database. There are also options for placing links to pop-ups with text and media files (images), basic data analysis/synthesis, intelligent data curation, etc. Aggregation of objects is done based on their common (one or more) characteristics depending on the specific model and application domain. The system summarizes groups of objects based on the aggregated data in order to improve the organization of object representation in subsequent analysis.

For each cultural object, all metadata is stored in the *media repository*. This metadata is represented in catalog records that point to the original media file(s) associated with each object. The *user profile repository* manages all user data and its changes.

For each cultural object, all metadata is stored in the *media repository*. This metadata is represented in catalog records that point to the original media file(s) associated with each object. The *user profile repository* manages all user data and its changes.

The *Administrative Services panel* offers mainly user data management, metadata export, tracking services, analysis services, etc.

User data management covers activities related to registration (including via OAuth technology - Sign up with Google, Facebook, etc.), data changes, setting the access level, etc.

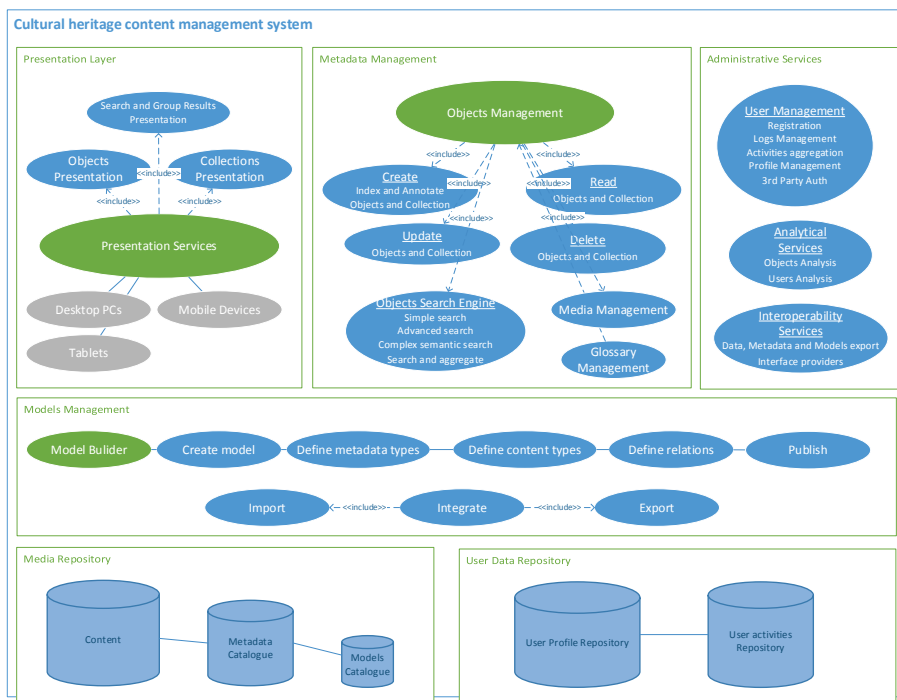


Fig. 1: Architecture of an environment for data storage, retrieval and curation of data in the field of humanities and social science

The *Administrative Services panel* offers mainly user data management, metadata export, tracking services, analysis services, etc.

User data management covers activities related to registration (including via OAuth technology - Sign up with Google, Facebook, etc.), data changes, setting the access level, etc.

Tracking services have two main branches: tracking access to the object's metadata and tracking user activity. The tracking of access to object's metadata basically monitors metadata management activities - adding, editing, visualizing, searching, deleting, selecting, exporting, grouping, etc. to provide a wide range of statistics (on service usage frequency, failed requests, etc.) for internal use and to generate inferences about the robustness (stability) and resilience of the operation and reliability of the environment. Tracking of the user's activities, monitors registrations, changes in personal data, access level and user behavior in the environment. It will support panels for visual analysis of the repository memory, based on online analytical processing and fast response to multi-dimensional analytical queries.

The technologies used for the development are as follows:

1. Back-end components:
 - Application layer: Node JS, Express 4;
 - Database: Mongo DB 4.4.
 - Additional (3^d party) application layer components:
 - Passport JS - OAuth 2 for Google & Facebook Authentication
 - FFMPEG & ImageMagick - for processing the digitized assets (water-marking and conversion to formats suitable for the web and relevant devices)
2. Back-end infrastructure:
 - OS Linux - Debian 10;
 - Node JS Server behind Apache 2.4 with Load Balancing and SSL;
 - Single Mongo DB node;
 - Hardware RAID 10 Storage for Media and DB repositories.
3. Front-end (web browser)
 - HTML, CSS (SCSS), Javascript (ES6);
 - SPA (Single Page Application), Vue JS (MVVM pattern), REST API based communication to the server side;
 - Bootstrap 4 to ensure responsiveness of the environment on different types of devices.
4. Git, WebPack - version control & release management.

3 Conclusions

Modern technology helps people to collaborate, work together, share knowledge and participate in social, political and cultural life, regardless of language barriers, cultural differences and computer skills. An important element in a knowledge sharing system is the existence of standardization and unification of processes and data. In this paper was presented the Bulgarian National Interdisciplinary Research E-Infrastructure for Bulgarian Language and Cultural Heritage Resources and Technologies and the developments of the Institute of Mathematics and Informatics at the Bulgarian Academy of Sciences that aimed to deliver and offer technologies and tools to unify the processes for access, preservation and use of cultural and historical content in accordance with established best practices and regulations in the field.

Acknowledgements

This work is funded in part 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 DO01-377/18.12.2020.

References

- 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, 2019, pp. 65-74, ISSN 1314- 4006 (Print), eISSN 2535-0366 (Online).
- Dochev, D., Pavlov, R., Paneva-Marinova, D., Pavlova, L. (2019), Towards Modeling of Digital Ecosystems for Cultural Heritage. *Digital Presentation and Preservation of Cultural and Scientific Heritage*. Vol. 9, Sofia, Bulgaria: Institute of Mathematics and Informatics - BAS, 2019, pp. 77-88, ISSN 1314-4006 (Print), eISSN 2535-0366 (Online).
- Spasova-Dikova, J. (2019), Transdisciplinarity and the New Ways for Preservation and Investigation of the Cultural and Historical Heritage. *Digital Presentation and Preservation of Cultural and Scientific Heritage*. Vol. 9, Sofia, Bulgaria: Institute of Mathematics and Informatics - BAS, 2019, pp. 89-100, ISSN 1314-4006 (Print), eISSN 2535-0366 (Online).
- Paneva-Marinova, D., Stoikov, J., Pavlova, L., Nikolova, A., Data Discovery and Distributed Representation for Better Cultural Heritage Observation and Learning. *Proceedings of the 13th annual International Conference of Education, Research and Innovation (ICERI2020)*, IATED, 2020, ISBN:978-84-09-24232-0, ISSN:2340-1095, 5696-5699, <https://librarv.iated.org/view/PANEVAMARINOVA2020DAT>

Received: June 02, 2021

Reviewed: July 05, 2021

Finally Accepted: July 15, 2021