

---

## PROPOSAL

for the establishment of a **full-time doctoral programme**  
pursuant to Art. 21(7) of the Higher Education Act

at the Institute of Mathematics and Informatics, Bulgarian Academy of Sciences  
Professional field: 4.6 Informatics and Computer Science  
Doctoral programme: *Informatics*

Dissertation topic:

### **Infrastructures and Open Solutions for Collections as Data in Bulgarian Cultural Heritage Institutions**

Proposed by: **Assoc. Prof. Milena Dobрева, PhD**

## Relevance

The Collections as Data paradigm represents a fundamental shift in the way cultural institutions organise their digital collections - not merely as objects for viewing and preservation, but as data that can be analysed, processed, and reused through computational methods. This paradigm shift imposes new requirements on institutions with regard to metadata quality, standardisation, legal frameworks for access and reuse, and the technical infrastructure for data publication.

In Bulgaria, GLAM institutions (galleries, libraries, archives, and museums) hold rich collections, yet their level of digital maturity varies significantly. Many maintain partially digitised catalogues with heterogeneous metadata formats, unclear or restrictive licensing regimes, and limited technical infrastructure for publishing open data. There is currently no comprehensive analysis of institutional readiness for transitioning to the Collections as Data model, nor a clear strategy to support such a transition.

At the European level, the regulatory framework provides favourable conditions - the Open Data Directive (EU 2019/1024), the Data Act (2023), the European Commission Recommendation on a Common European Data Space for Cultural Heritage (2021), as well as initiatives such as EOSC, the Europeana Data Space for Cultural Heritage, and the European Open Science Cloud. However, their implementation at the institutional level in Bulgaria remains fragmented and insufficiently studied. At the same time, the Wiki ecosystem (Wikidata, Wikimedia Commons) is increasingly established as a de facto infrastructure for linked open data in cultural heritage, while national and European open data portals (e.g. data.government.bg, data.europa.eu) provide publication platforms that remain underutilised by GLAM institutions.

Meanwhile, the growing application of artificial intelligence (AI) in the cultural sector introduces new requirements regarding the quality, structure, and accessibility of institutional data. AI functions both as a tool for data processing and enrichment (automatic classification, object recognition, metadata generation) and as a primary consumer of well-structured linked open data - making institutional readiness for data openness an increasingly urgent issue.

The proposed doctoral research is directly aligned with the priorities of the FOCUS project, particularly the areas of “Datafication of Cultural Heritage” and “Open Knowledge Infrastructures”. While other strands of the project focus on technical aspects of data integration and enrichment,

this doctoral research addresses the institutional, legal, and organisational dimensions—without which technical solutions cannot be sustainably implemented.

## Objective of the Doctoral Research

The aim of the doctoral research is to develop a methodological framework for assessing the readiness of GLAM institutions to open and publish cultural heritage data, as well as to design experimental technological solutions that support this transition and formulate a roadmap towards the Collections as Data model.

## Research Tasks and Methods

- Development of an institutional readiness assessment model (maturity model), including dimensions such as metadata quality and completeness, standards used, level of digitisation, technical infrastructure, organisational capacity, and access policies;
- Assessment of data quality in selected institutions through application of the model, including analysis of completeness, consistency, granularity, interoperability, and compliance with FAIR principles;
- Identification of barriers and enablers for opening institutional data—technical, organisational, legal, and cultural;
- Investigation of the integration with the Wiki ecosystem (Wikidata, Wikimedia Commons) as a strategy for increasing openness and visibility, including analysis of best practices from European GLAM institutions;
- Evaluation of readiness for publication via national and European open data portals ([data.government.bg](http://data.government.bg), [data.europa.eu](http://data.europa.eu)), including analysis of technical and organisational prerequisites for sustainable publishing;
- Analysis of the prerequisites for the application of AI in GLAM institutions, with a focus on data quality and completeness as conditions for effective use of AI tools;
- Validation of the framework through pilot implementation in selected institutions, with stakeholder involvement;
- Selection and, where necessary, development of supporting software tools facilitating the transition from digital collections to Collections as Data.

## Expected Results

The research is expected to deliver:

- a maturity model for assessing institutional readiness for Collections as Data, applicable across institutions of different types and scales, including legal aspects;
- empirical data from the assessment of specific institutions, demonstrating the applicability of the model;
- a transition model for institutions with existing digital collections towards the Collections as Data paradigm;
- analysis of opportunities for integration with the Wiki ecosystem and open data portals as channels for publication and visibility;
- software tools supporting cultural institutions in the transition to Collections as Data.

## Impact

The expected results will contribute directly to strengthening the institutional capacity for publishing open cultural heritage data in Bulgaria. The maturity model and roadmap will provide practical tools for planning and implementing the transition to Collections as Data.

At the policy level, the analysis of the legal framework and identified barriers will support the development of more effective open data policies in the cultural sector, in alignment with European initiatives for a Common European Data Space for Cultural Heritage.

From an informatics perspective, the maturity model and data quality metrics have the potential for application in other domains where institutional data undergo processes of opening and transformation.

More broadly, the results will support the integration of Bulgarian cultural heritage into European and international infrastructures (e.g. EOSC, Europeana, Wikidata) and promote the adoption of Open Science principles in the humanities and social sciences. The research will also contribute to preparing institutions for the effective use of AI technologies by ensuring the availability of high-quality, standardised, and accessible data without which the potential of AI in the cultural sector cannot be fully realised.

## Affiliation

Department of Software Technologies and Information Systems (Institute of Mathematics and Informatics, BAS)

## Used Research Infrastructure

- CLADA-BG
- HEMUS Supercomputer