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:

Methods and Technologies for Integration, Semantic Enrichment, and Mapping of Open Data for Public Art

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

Relevance

Public art constitutes a significant yet insufficiently documented and systematically studied domain of cultural heritage. It encompasses diverse forms – sculptures, murals, installations, temporary artistic interventions, commemorative plaques – closely embedded in specific urban, rural, environmental, and social contexts. Data on public art are dispersed across numerous heterogeneous sources: institutional registers, scholarly publications, digital archives, open knowledge platforms (e.g. Wikidata, OpenStreetMap), and increasingly social media and informal online communities.

This fragmentation leads to substantial challenges related to data incompleteness, inconsistency, and lack of standardisation. In many cases, information is partial, duplicated, or difficult to reconcile, which limits its usability and reuse. The absence of widely adopted descriptive models, as well as effective mechanisms for integrating data from diverse sources, constrains the ability to construct a comprehensive and reliable representation of public art as part of the cultural ecosystem.

From an informatics perspective, this raises a range of challenges related to data modelling, integration of heterogeneous sources, semantic enrichment, and automated knowledge extraction. There is a clear need for methods that can transform fragmented information into structured and interconnected datasets compliant with the principles of open data. The lack of such solutions restricts the use of public art data in various applications, including research, education, tourism, the creative industries, and urban planning. At the same time, there is growing interest at the European level in developing open and linked cultural heritage data, including within initiatives such as the European Collaborative Cloud for Cultural Heritage and the European Open Science Cloud (EOSC).

The proposed doctoral research is directly aligned with the priorities of the FOCUS project, particularly the areas of “Datafication of Cultural Heritage,” “AI in GLAM,” and “Open Knowledge Infrastructures.” By developing methods for integrating and enriching public art data, the research contributes to transforming cultural objects into analysable and reusable digital resources, thereby supporting the development of sustainable research and innovation ecosystems in digital cultural heritage.

Objective of the Doctoral Research

The aim of the doctoral research is to develop and experimentally validate computational methods and models for the integration, semantic enrichment, and analysis of open data on public art.

To achieve this aim, the following specific objectives are defined:

- **Identification and systematisation of data sources on public art in Bulgaria**, potentially through two case studies (e.g. an urban district in the capital and a small town), including analysis of legal and licensing conditions for use and reuse;
- Definition of a conceptual and semantic model ("object passport") for public art objects, and investigation of possibilities for its (semi-)automatic generation through integration of open and distributed data;
- **Development and experimentation with methods for integrating heterogeneous data**, including approaches for alignment, enrichment, and linking of information from diverse sources;
- **Investigation of methods for spatial analysis, mapping, and visualisation of public art data**, including the integration of geospatial and semantic representations;
- **Specification of the requirements for a prototype software environment** integrating data discovery, extraction, modelling, integration, and visualisation processes;
- **Analysis and validation through user scenarios targeting different stakeholder groups** – researchers, GLAM institutions, the public sector, and the creative industries.

Research Tasks and Methods

The research will begin with a comprehensive analysis and evaluation of existing data sources on public art and the gaps within them. This includes identifying relevant sources, assessing their quality, completeness, and compatibility, and analysing issues related to fragmentation, duplication, and lack of standardisation. Specific challenges related to the curation of such artefacts will also be examined.

Based on this analysis, a conceptual and semantic data model will be developed to enable structured representation of public art objects, their attributes, relationships, and context. Particular attention will be given to defining the "object passport" as a formalised unit for description, integration, and data exchange.

The next phase will focus on the development and experimental evaluation of methods for integrating heterogeneous data from multiple sources. This will include approaches to entity matching and linking, deduplication, and harmonisation of data structures and metadata, with the aim of constructing consistent and interconnected datasets. Methods for spatial analysis, mapping, and visualisation will also be explored, enabling interpretation of data within their geographical and temporal context.

Approaches for combining geospatial and semantic representations, as well as for developing interactive visual environments, will also be investigated.

Expected Results

The research is expected to result in the development of new computational methods and models for the integration, semantic enrichment, and analysis of heterogeneous public art data. Specifically, the following outcomes are anticipated:

- Methods for integration and alignment of data from diverse open and institutional sources, including approaches for entity recognition, linking, and deduplication;
- A conceptual and semantic model ("object passport") enabling standardised description, exchange, and reuse of public art data;
- Methods for combined spatial and semantic analysis supporting the study of public art in its geographical, temporal, and social context;
- Formalised user scenarios for the application of integrated data in research, education, cultural policy, and the creative industries;
- Guidelines for the development of a prototype software system implementing the proposed methods and enabling experimental validation with real-world data;
- Scientific publications presenting results in the fields of informatics, digital cultural heritage, and data analysis.

Impact

The expected results will contribute to the advancement of digital cultural heritage by introducing new approaches to the structuring, integration, and analysis of public art data. The project will support data reuse and facilitate access for diverse user groups, including researchers, cultural institutions, public administrations, and citizens.

The proposed methods and models will enable the integration of Bulgarian cultural heritage into European and international infrastructures such as EOSC and Europeana, and will support the implementation of Open Science and Collections as Data principles.

From an informatics perspective, the project will contribute to the development of methods for working with heterogeneous and distributed data, with potential applications beyond the cultural heritage domain.

More broadly, the results will support the development of tools for cultural planning, tourism, and the creative industries, as well as enhance understanding and dissemination of public art as an integral part of the contemporary cultural environment.

Affiliation

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

Used Research Infrastructure

- CLADA-BG
- HEMUS Supercomputer