Museum Big Data: Emerging Practices, Challenges, and Opportunities in Museum Contexts – A Preface and a Synthesis of the MBD2024 Conference

Georgios Papaioannou[1,*,†] and Matthew Damigos [1,†]

[1] *Ionian University, Dept. of Information Science, Corfu 49100, Greece*

Abstract

The *3rd International Conference on Museum Big Data (MBD2024)*, held in Athens, Greece, in November 2024, brought together researchers, professionals, and digital innovators to explore the intersections between big data technologies and museum practice. The conference showcased a wide range of initiatives employing linked data, Artificial Intelligence (AI), immersive visualizations, and semantic modeling to enhance cultural heritage documentation, interpretation, and engagement. Key aspects related to integration of knowledge graphs and ontologies in museum contexts, the role of immersive and participatory technologies in public engagement, and the ethical challenges posed by data collection in GLAM institutions. This paper provides a synthesis of the conference sessions, highlighting current trends, methodological advances, and future directions for research and innovation in digital museology and data-driven heritage.

Keywords

Museum Big Data, Big Data, Museums, Digital Heritage, Linked Open Data, Knowledge Graphs, CIDOC CRM, Immersive Technologies, Artificial Intelligence, GLAMs, Ontologies, Cultural Informatics, Museum Informatics, Ethics and Data Governance

1. Introduction

The rapid growth of digital technologies has fundamentally transformed the ways in which museums and cultural heritage institutions collect, preserve, and disseminate information [1, 2]. The proliferation of digitized collections [3], user-generated content, and sensor-derived data has led to the emergence of "big data" in the cultural sector—large, complex, and heterogeneous datasets that offer unprecedented opportunities for knowledge creation and audience engagement. When properly integrated and analyzed, Big Data enables museums to uncover previously hidden relationships among collections, to provide richer and more personalized visitor experiences [4], and to inform strategic decision-making through evidence-based insights. At the same time, the use of Big Data in cultural environments raises significant challenges related to interoperability, data quality, privacy, use of Artificial Intelligence [5] and ethical stewardship. Addressing these issues is essential to realizing the transformative potential of big data for enhancing both the scholarly understanding of cultural heritage and its social impact.

Within this context, the 3rd International Conference on Museum Big Data (MBD2024) convened at the National and Kapodistrian University of Athens, Greece, on 18–19 November 2024. Organized jointly by the Ionian University (Museology Research Lab) and the National and Kapodistrian University of Athens (Electronic Management of Historical Archives Lab), MBD2024 brought together academics, museum professionals, big-data experts, and digital innovators to address these issues, to reflect and discuss, and to direct and project future developments in this emerging field of study, research and practice.

MBS2024: 3rd International Conference On Museum Big Data, November 18-19, 2024, Athens, Greece

* Corresponding author.

† These authors contributed equally.

gpapaioa@ionio.gr (G. Papaioannou); mgdamigos@ionio.gr (M. Damigos)

0000-0002-9270-0463 (G. Papaioannou); 0000-0003-0432-1482 (M. Damigos)

This paper offers an overview of the conference proceedings, outlining emerging trends and methodological developments, as well as prospective avenues for research and practice within museology, museum practices and data-informed cultural heritage issues. It consists of the following sections: an introduction (section 1), a brief history of the Museum Big Data Conferences and their scope and vision (section 2), an overview of the MBD2024 proceedings (section 3), a reflection/synthesis of the conference works (section 4), and concluding and future work remarks (section 5).

This paper also serves as the Preface of papers published in this CEUR-WS volume. There were 19 papers submitted for peer-review to MBD2024. Out of these, 16 papers were accepted for this volume, all of them as regular papers. We had also five invited papers (four keynotes and a session paper). The session invited paper is from Dr Sofia Chatzidi and it is published as a short paper.

2. The Museum Big Data Conferences: Scope, Vision and a brief History

The Museum Big Data (MBD) conference series (https://museumbigdata.org/) constitutes a dynamic and evolving platform for the dissemination and exchange of high-quality research, experimental practice, and interdisciplinary dialogue on the intersection of big data and the cultural heritage domain. Since its inception in 2019, MBD has consistently sought to foreground critical discussions at the convergence of museum studies, data science, and digital cultural heritage, attracting a diverse audience of scholars, professionals, and technologists from the broader GLAM sector (Galleries, Libraries, Archives, and Museums).

The primary objective of MBD is to cultivate a research-intensive environment that not only supports rigorous academic inquiry but also fosters collaboration between institutions, disciplines, and professional communities. Moreover, MBD conference series aim to shape the research agenda in the domain of Museum Big Data, identifying challenges and promoting solutions through reflection, discussion, evaluation, and decision-making processes. By highlighting ongoing projects, early-stage innovations, and high-impact case studies, MBD conferences aims to serve as catalysts and incubators for future scholarly and institutional developments in the museum sector.

Big data and relevant processes (including data mining, predictive modeling, and pattern detection) have emerged as central areas of practice and research within the world of museums. As museums and cultural institutions worldwide continue to digitize collections and accumulate data from a wide range of activities, including museum object documentation, visitor analytics, and social media engagement, new opportunities and challenges have arisen in terms of interpretation, management, governance, use and reuse of these data. Consequently (and inevitably), the development of computational approaches producing and identifying hidden patterns, relationships, and insights is becoming essential for museums seeking (even obliged) to operate in data-rich and data-driven environments.

In response to this landscape, the MBD conferences highlight the necessity of interdisciplinary approaches integrating ethical considerations, contextual sensitivity, and technological expertise. The aim is to bridge the physical and digital dimensions of museum practices and operations, while also participating, exchanging and encouraging collaborations with similar institutions and organizations, such as libraries and information centers, galleries and archives, within the emerging world of digital heritage and humanities. The integration of theoretical models, comparative frameworks, and cross-institutional practices will be critical for ensuring that big data initiatives in museums not only enhance operational efficiency but also enrich scholarly understanding and public engagement.

The MBD conference series began with its inaugural event (MBD2019) held from April 30 to May 2, 2019, in Doha, Qatar, thanks to a generous research grant by Qatar Foundation to Prof Georgios Papaioannou, envisioning the establishment of an academic forum for research exchange and international collaboration in the emerging field of Museum Big Data. MBD2019 brought together

leading international scholars and practitioners to present and explore how museums could better adopt and integrate Big Data practices into their curatorial, educational, and managerial activities, policies and practices. The 2nd MBD was held in Nicosia, Cyprus, at the premises of The Cyprus Institute in 2020 (MBD2020), continuing the series and establishing the MBD scope and vision.

The MBD2024 conference, the 3rd International Conference on Museum Big Data, continues the MBD mission and vision, offering a transdisciplinary platform for scientific exchange, reflection and experimentation on data science and museum practices. Given the fact that the volume, the variety and the complexity of museum-related data continue to rapidly grow, so does the need for critical infrastructures, interoperable frameworks, and participatory models supporting the ethical, sustainable, and meaningful use of such data. To address these, MBD offers the floor and enhances open dialogue, methodological innovation, and a commitment to scholarly excellence. The MBD Conference series is scheduled to continue in 2026, with the next conference (MBD2026) planned to take place in Brazil.

For MBD2024, we thank Prof Milena Dobreva, Honorary Chair of MBD2024, who has been involved in the MBD conferences from the very first moment of MBD2019 onwards; Profs Christos Papatheodorou and Michalis Sfakakis, members of the Program Committee; our keynote speakers (alphabetically) Erik Champion, Jill Cousins, Laurent D' Orazio, Panos Konstantopoulos; all members of the MBD2024 Organizing and Scientific Committees; all speakers and participants.

3. The MBD2024 Proceedings: An Overview

After a welcome by Prof. Georgios Papaioannou, Head of MBD Conferences, three keynote presentations supported MBD2024' scope and vision, offering reflection starting posts: Constantopoulos delivered "Random thoughts on big data in museums", setting the stage for conceptual reflection on museum data paradigms; Champion explored "Immersive Visualisation and the Emergence of Collaborative XR in the Museum Sector", focusing on extended reality technologies for participatory museum experiences; and d'Orazio analyzed "Big Data in museums: a brief history of cloud data management and perspectives", underscoring strategic and infrastructural implications for GLAM institutions.

The MBD2024 conference was structured around six thematically interconnected sessions, each reflecting critical areas of research and practice within the evolving domain of Big Data in museums and cultural heritage. The first session entitled "Documentation & Semantics", addressed the theoretical and technical underpinnings of data representation, focusing on semantic modeling, ontologies, and knowledge organization systems. Presentations explored the application of formal semantic frameworks such as CIDOC CRM, the integration of controlled vocabularies, and the development of interoperable documentation infrastructures that enhance data richness and accessibility. Angelaki et al. presented the SearchCulture.gr (the Greek National Cultural Heritage Aggregator), leveraging Linked Data to unify heterogeneous national cultural data; Avgousti, Papaioannou & Koutoupas discussed digital accessibility methods for Big Data in heritage contexts; Nikolaidou introduced an ontology-based framework for documenting artists' studios, emphasizing semantic representation

The second session was on "Content Management, Data Collection & Curation" and focused on strategies for gathering, managing, and curating cultural data at scale. Contributions discussed novel methodologies for data acquisition from diverse sources, the integration of heterogeneous datasets, and curatorial practices that support sustainable and ethically informed data management. This section also emphasized the importance of metadata quality, standards compliance, and institutional workflows in shaping usable and meaningful data repositories. Koutoupas et al. showcased ARTES, a digital twin-driven platform for Big Data -infused art management; Chagas et al. proposed a curated model linking local history and museum data for heritage education in Brazil; Avgousti, Koutoupas & Bakirtzis analyzed digital archives from Pancyprian Gymnasium in Cyprus, demonstrating the use of Big Data tools in archival context.

In the third session "From Small Places to Big Data: Case Studies", case studies were presented, offering insights into how smaller institutions or community-driven initiatives can effectively contribute to (and benefit from) Big Data through tailored methodologies and context-sensitive innovations. Sfyridou examined how small cultural institutions in Cyprus deploy websites and social media effectively. Andrianou explored data-driven interpretation of cultural landscapes via the martyr village of Kommeno in Greece; Chalkia, Douka & Sfyridou discussed how Big Data influences digital literacy, interpretative reading, and writing skills.

The fourth session, "AI and Big Data", examined the intersection of artificial intelligence with cultural data processing. Presenters explored applications of machine learning, natural language processing, and predictive analytics in the museum domain, reflecting on both the opportunities and the epistemological challenges posed by AI-enhanced interpretation and automation in heritage work, including ethical issues and personal and cultural data handling. Mountantonakis, Koumakis & Tzitzikas presented a case study on combining large language models and hundreds of Knowledge Graphs to enrich and validate cultural heritage data; Deliyannis et al. discussed personalized tourist experiences using real-time data; Pediaditaki addressed GDPR and ethical challenges in handling personal and cultural data.

The fifth session was on "Visualising Data and Collections" and brought together contributions that interrogated the role of visualization as both a research method and a tool for public engagement and policy making. Projects presented in this session illustrated innovative uses of dashboards, digital cartographies, immersive experiences, interactive data storytelling, and national/international heritage policies to render complex cultural datasets intelligible and compelling to diverse audiences. Artopoulos, Loucas & Daune-Le Brun introduced an immersive workflow for participatory reconstruction of archaeological sites; Chagas, Filgueiras & Gouveia detailed educational reuse of museum collections in schools in Brazilian regional towns; Papadopoulou shared a conceptual Folk Tale Museum scenario in Zagori, Greece, employing Big Data for exhibit design; Chatzidi presented contrasting national and international museum policies shaping the global trade in cultural objects.

Finally, the sixth session on "Communities, Country Groups & GLAMs" foregrounded participatory and inclusive approaches to Big Data in cultural heritage. It focused on collaborations across galleries, libraries, archives, and museums (GLAMs), as well as community-based initiatives and transnational efforts to foster equitable access to data, representation in collections, and cross-sectoral dialogue. Dobreva & Papaioannou discussed the GLAM Lab, an innovation sprint culminating in a sprint-book; Aggeletaki & Mavroudi focused on how higher education-driven open innovation aids cultural heritage recovery.

Together, these six sections constituted a comprehensive exploration of current trends, critical reflections, and emerging directions in Museum Big Data, reaffirming the field's commitment to interdisciplinarity, innovation, and ethical engagement.

The final session consisting of keynote speech and a panel discussion by Cousins, founder of Europeana, underlined the importance and necessity of a shared unified vision for open data in cultural heritage institutions, stressing collaboration, standards, and interoperability across the GLAM sector.

4. The MBD2024: Reflection and Synthesis

The MBD2024 conference (18-19 November 2024) served as a comprehensive and dynamic forum for scholarly and professional discussion and reflection in the emerging field of Museum Big Data. A key achievement of MBD2024 was bringing together diverse perspectives (i.e. research, practice, policy, and design) into a cohesive and well-organized dialogue. All six sessions emphasized how semantic technologies, data science methods, and user/visitor-centered design can influence the ways in which museums and cultural institutions create, manage, and share content and knowledge in the 21st century. From digital twins and linked data ontologies to participatory platforms and

immersive storytelling, MBD2024 papers discussed and critically reflected on Museum Big Data aspects, conforming that it is a field that becomes increasingly central in the museum world.

In terms of themes that have emerged as central across the six conference sections, the keynote speeches and the discussion, we highlight the foundational role of Linked Open Data, Knowledge Graphs, and semantic standards. It has emerged as essential, particularly for enhancing interoperability across platforms, as it enables multiple and meaningful connections among datasets and long-term sustainability in digital museum and cultural records, ensuring that museum and heritage data infrastructures be (and remain) open, scalable, and adaptable to changing technological, institutional, and social needs and realities. Another key aspect was the area or immersive technologies, including XR (Extended Reality), AR (Augmented Reality), and VR (Virtual Reality), in visitor-related applications. These technologies and their importance have been addressed both as instruments for museum display (either supporting object(s) exhibition or offering digital / multimedia exhibits themselves), but also as tools for audience engagement, enabling new forms of interpretation, connection, and cultural literacy. Ethical considerations, including data policies, personal data, privacy, data protection regulations (such as GDPR), transparency, and accountability have emerged as Museum Big Data issues in museum practices, stressing the principal need to deal with ethical reasoning as an integral part of design and implementation, rather than an imposed addendum. MBD2024 also underscored the need for solid bridges between research work and museum/institutional practice. Papers addressed museum current practices in need of implementing data models and AI tools within their operation, acknowledging that vision and reality do not necessarily (and definitely not always) go together in museum (and in GLAM) contexts due to limitations related to funding, expertise and resource availability.

In sum, MBD2024 managed to present the latest scholarly work and initiatives as well as to offer a space for critical discussion, reflection and synthesis, highlighting challenges, proposing solutions, and fostering a sense of collective purpose in the rapidly evolving field of Museum Big Data.

5. Conclusions and Future Work

Starting from future work, MBD2024 pointed to key directions expected to influence the future of research and practice within museums and Museum Big Data. A priority is to locate and study case studies of museums and cultural institutions (smaller and bigger) that have incorporated Museum Big Data applications, policies, tools and strategies. Getting to know, discuss, reflect upon, and synthesize based on real-life examples will help towards addressing issues and develop more effective approaches. Secondly, the integration of artificial intelligence (AI), especially multimodal and generative AI, represents an emerging frontier. The convergence of large language models, vision-language models, and structured data sources (e.g., Knowledge Graphs) opens new possibilities for enhancing search, classification, storytelling, and visitor personalization in museums, provided that clear methodological frameworks are implemented and rigorous critical analysis are applied. Co-designed and co-created tools, such as user-centered designed platforms, metadata automated systems, curation interfaces and dashboards, will bridge the gap between advanced computational capabilities and real-world museum needs. To this end, strengthening of collaborative infrastructures, such as ARTES and SearchCulture.gr, will help. These platforms enable cross-institutional data exchange, facilitate standardization efforts, and serve democratization and innovation, shaping a more connected and interoperable museum (and data) environment. Finally, effort and investment are needed in ethical and legal literacy, including data-related training for museum professionals. As museums engage more deeply with complex data environments, which themselves become more and more complex and elaborated, museums and cultural institutions must address responsibilities that inevitably come along, related to both their publics and visitors, and to the data / cultural records they preserve, use, and curate.

In conclusion, MBD2024 brought together scholars and practitioners, contributing to new partnerships, up-to-date discussions, and shared understandings. By providing critical reflection on the state of the field, it identified issues, strengths and areas for further action, research, and

development. The publication of the MBD2024 proceedings in the CEUR Workshop Series ensures that the MBD2024 works will continue to inform museum communities on Museum Big data -related research and practices. As the digital transformation of museums and cultural heritage becomes stronger and accelerates, the MBD conference series will continue to meaningfully contribute to the evolving field of Museum Big Data and data-driven museology, with a sustained focus on ethical, inclusive, and culturally sensitive approaches to museum innovation, research and practices.

Acknowledgements

The MBD2024 was supported by EasyConferences (https://easyconferences.eu/) and the EasyChair patlform (https://easychair.org/). The online aspect was supported by the Networks Operation Centre of the Ionian University, Greece.

Declaration on Generative AI

The authors have not employed any Generative AI tools.

References

[1] F. Amato, V. Moscato, A. Picariello, F. Colace, M.D. Santo, F.A. Schreiber, I. Tanca, Big data meets digital cultural heritage: Design and implementation of scrabs, a smart context-aware browsing assistant for cultural environments. Journal on Computing and Cultural Heritage 10/1 (2017) 1-23.

[2] G. Papaioannou, Museum big data: perceptions and practices, in: Th. Prodromou (Ed.), Big Data in Education: Pedagogy and Research, 2021, pp. 201-215.

[3] Y. Yu, Y. (2020, October). Big data technology in museum exhibition digitization, Journal of Physics: Conference Series 1648/4 (2020, October), p. 042044.

[4] C.A. Dimoulas, Cultural heritage storytelling, engagement and management in the era of big data and the semantic web, Sustainability 14/2 (2022) 812.

[5] G. Buratti, S. Conte, M. Rossi, Artificial intelligence, big data and cultural heritage, DISÉGNO-OPEN ACCESS (2021) 29-34.