

IONIAN UNIVERSITY



FACULTY OF INFORMATION SCIENCE & INFORMATICS

DEPARTMENT OF ARCHIVES, LIBRARY SCIENCE

AND MUSEOLOGY

Master of Science in

«Digital Information Management - Information Services»

MASTER'S THESIS

«How big is Big? The Data of a Major Museum: Their Volume,
Management, and Future Prospects (Case Study: The British
Museum in London)»

Anastasia Zika

Corfu September, 2025

This page is intentionally blank

IONIAN UNIVERSITY
FACULTY OF INFORMATION SCIENCE & INFORMATICS
DEPARTMENT OF ARCHIVES, LIBRARY SCIENCE AND
MUSEOLOGY

Master of Science in
«Digital Information Management - Information Services»

MASTER'S THESIS
«How big is Big? The Data of a Major Museum: Their Volume,
Management, and Future Prospects (Case Study: The British Museum in
London)»

Anastasia Zika
Student ID: alm.dimis2305


Supervisor:
Georgios Papaioannou, Department of Archives, Library Science and
Museology, Professor

Master's Thesis Examining Committee:
Georgios Papaioannou, Department of Archives, Library Science and
Museology, Professor
Matthew Damigos, Ionian University Faculty of Information Science &
Informatics, Department of Archives, Library Science and Museology,
Assistant Professor
Eleftherios Kalogeros, Department of Archives, Library Science and
Museology, Laboratory Teaching Personnel

Corfu September, 2025

AUTHOR'S DECLARATION:

I explicitly declare that, in accordance with Article 8 of the Greek Law 1599/1986, and Articles 2, 4, and 6, paragraph 3 of the Greek Law 1256/1982, the present work is exclusively the result of my personal effort and does not infringe on any form of third-party intellectual property rights. It is not a product of partial or complete plagiarism, and the sources used are limited solely to the bibliographical references.

Author's signature: 

This Thesis and its results are the joint property of the Ionian University and the author. Each of these two parties has the right to independently use, reproduce and redistribute the thesis and/or its results (in whole or in part) for educational and research purposes, always citing the title and author of the Thesis, together with the name of the Ionian University where the Thesis has been presented and examined.

Acknowledgements / Dedication

I would like to express my gratitude to the Ionian University, and in particular to the Department of Archives, Library Science and Museology of the Faculty of Information Science and Informatics, for their guidance and support throughout my undergraduate and postgraduate studies. The knowledge and skills I gained during my time here have been invaluable to both my academic growth and professional development.

This page is intentionally blank

Abstract

This thesis explores the role of big data in museums, with a focus on the British Museum as a case study. Modern museums are no longer simply custodians of physical artifacts; they have evolved into complex data-driven institutions that generate, manage, and interpret vast amounts of digital information. This includes digitized collections, metadata, high-resolution images, conservation records, visitor analytics, educational resources, and user-generated content. The research investigates how such data is collected, curated, stored, and strategically used, and examines both the opportunities and challenges that arise in this process.

The British Museum, with more than 8 million objects and millions of digitized records, represents a prime example of a large-scale museum grappling with the realities of big data. The study highlights how digital infrastructures such as collection management systems, APIs, linked open data, and immersive technologies contribute to accessibility, interoperability, and public engagement. At the same time, it considers the challenges posed by data heterogeneity, privacy concerns under GDPR, and the sustainability of large digital ecosystems.

Key findings indicate that while big data allows museums to expand access, personalize visitor experiences, and foster educational innovation, it also requires strong governance frameworks, ethical safeguards, and sustainable infrastructures. The thesis contributes to current discussions on digital curation, data ethics, and visitor analytics, offering insights into how major cultural institutions balance openness, engagement, and responsibility in the digital age.

By focusing on the British Museum as a leading example, this research demonstrates how cultural heritage institutions can strategically manage big data to fulfill their mission of preservation, education, and accessibility, while also preparing for future challenges posed by emerging technologies such as artificial intelligence and virtual reality.

Keywords:

Big Data, Museums, Digital Curation, British Museum, Data Governance, Cultural Heritage, Visitor Analytics, Open Access

This page is intentionally blank

«Πόσο Μεγάλο είναι το “Μεγάλο”; Τα δεδομένα ενός μεγάλου μουσείου: ο όγκος τους, η διαχείριση και οι μελλοντικές τους προοπτικές (Μελέτη περίπτωσης: Το Βρετανικό Μουσείο στο Λονδίνο)»

Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει τον ρόλο των Big Data στα μουσεία, με μελέτη περίπτωσης το Βρετανικό Μουσείο. Τα σύγχρονα μουσεία δεν αποτελούν πλέον απλούς φύλακες υλικών αντικειμένων· έχουν εξελιχθεί σε σύνθετους, ψηφιακά καθοδηγούμενους οργανισμούς που παράγουν, διαχειρίζονται και ερμηνεύουν τεράστιους όγκους ψηφιακών δεδομένων. Αυτά περιλαμβάνουν ψηφιοποιημένες συλλογές, μεταδεδομένα, εικόνες υψηλής ανάλυσης, εκθέσεις συντήρησης, αναλύσεις επισκεπτών, εκπαιδευτικό υλικό και περιεχόμενο που δημιουργείται από τους ίδιους τους χρήστες. Η έρευνα διερευνά πώς αυτά τα δεδομένα συλλέγονται, επιμελούνται, αποθηκεύονται και αξιοποιούνται στρατηγικά, ενώ εξετάζει παράλληλα τις ευκαιρίες αλλά και τις προκλήσεις που προκύπτουν.

Το Βρετανικό Μουσείο, με περισσότερα από 8 εκατομμύρια αντικείμενα και εκατομμύρια ψηφιοποιημένες εγγραφές, αποτελεί χαρακτηριστικό παράδειγμα ενός μεγάλου μουσείου που αντιμετωπίζει τις πραγματικότητες των Big Data. Η μελέτη αναδεικνύει πώς οι ψηφιακές υποδομές, όπως τα συστήματα διαχείρισης συλλογών, τα APIs, τα Linked Open Data και οι τεχνολογίες εμπύθισης, συμβάλλουν στην πρόσβαση, τη διαλειτουργικότητα και τη συμμετοχή του κοινού. Την ίδια στιγμή, εντοπίζονται προκλήσεις όπως η ετερογένεια των δεδομένων, τα ζητήματα ιδιωτικότητας με βάση τον GDPR και η βιωσιμότητα των μεγάλων ψηφιακών οικοσυστημάτων.

Τα βασικά συμπεράσματα δείχνουν ότι, παρότι τα Big Data επιτρέπουν στα μουσεία να διευρύνουν την πρόσβαση, να εξατομικεύουν τις εμπειρίες των επισκεπτών και να προωθούν την εκπαιδευτική καινοτομία, απαιτούν επίσης ισχυρά πλαίσια διακυβέρνησης, ηθικές εγγυήσεις και βιώσιμες υποδομές. Η εργασία συμβάλλει στον

τρέχοντα διάλογο για την ψηφιακή επιμέλεια, την ηθική των δεδομένων και την ανάλυση επισκεπτών, προσφέροντας εικόνα για το πώς μεγάλοι πολιτιστικοί οργανισμοί ισορροπούν ανάμεσα στη διαφάνεια, τη συμμετοχή και την ευθύνη στην ψηφιακή εποχή.

Εστιάζοντας στο Βρετανικό Μουσείο ως ηγετικό παράδειγμα, η έρευνα δείχνει πώς τα πολιτιστικά ιδρύματα μπορούν να διαχειριστούν στρατηγικά τα Big Data ώστε να εκπληρώνουν την αποστολή τους για διατήρηση, εκπαίδευση και προσβασιμότητα, ενώ ταυτόχρονα προετοιμάζονται για τις μελλοντικές προκλήσεις που φέρνουν τεχνολογίες όπως η τεχνητή νοημοσύνη και η εικονική πραγματικότητα.

Λέξεις-κλειδιά:

Big Data, Μουσεία, Ψηφιακή Επιμέλεια, Βρετανικό Μουσείο, Διακυβέρνηση Δεδομένων, Πολιτιστική Κληρονομιά, Ανάλυση Επισκεπτών, Ανοικτή Πρόσβαση

This page is intentionally blank

Table of Contents

Table of Contents	vii
List of Figures	x
List of Tables	x
List of Abbreviations	xi
Chapter 1 Introduction.....	13
1.1 Background and Scope of the Study	13
1.1.1 The Data-Driven Museum Landscape	14
1.1.2 Why the British Museum? Justification of the Case Study	15
1.1.3 Research Aims and Objectives	16
1.1.4 Research Questions	17
1.2 Figures and Graphs	18
1.3 Tables	20
Chapter 2 Literature Review	23
2.1 Museums in the age of Big Data.....	23
2.1.1 The Shift to Data-Driven Museums.....	23
2.1.2 Opportunities of Big Data in Museums	24
2.1.3 Challenges of Big Data in Museums	25
2.2 Digital Curation and Data Governance in Cultural Institutions	26
2.2.1 Digital Curation in the Museum Ecosystem	26
2.2.2 Standards and Frameworks for Digital Curation.....	27

2.2.3 Platforms and Systems for Managing Museum Data.....	27
2.2.4 Data Governance and Policy in Cultural Heritage	28
2.2.5 Emerging Approaches: Educational and Immersive Curation	28
2.3 Open Access, Data Ethics and Visitor Analytics.....	29
2.3.1 Open Access and Ethical Responsibility.....	29
2.3.2 GDPR, Privacy, and Consent in Digital Collections	30
2.3.3 Analytics, AI, and Visitor Personalization	31
2.3.4 Open Cultural Data: Education, Participation and Interpretation	32
Chapter 3 The British Museum: Data Ecosystem.....	35
3.1 Types and Sources of Data	35
3.1.1 Collection & Object Data.....	35
3.1.2 Visitor & Operational Data.....	37
3.1.3 Digital Engagement & Media.....	39
3.1.4 Research & Collaborative Data	41
3.2 Digital Infrastructure and Platforms at the British Museum	44
3.2.1 Online Collection Platforms and Data Access	44
3.2.2 Data Standards and Interoperability.....	45
3.2.3 Immersive Technologies and Virtual Reality.....	47
3.3 Scale of the British Museum’s Collection Data (Comparative Perspective)	49
Chapter 4 Data Management and Strategic Use.....	52
4.1 Data Storage, Access, and Preservation Strategies	52
4.2 Using Data for Engagement, Personalization, and Learning.....	54

4.2.1 Digital Exhibits / Interactive Experiences.....	55
4.2.2 Visitor Personalization / AI	56
4.2.3 Educational Applications / Digital Tools	58
4.3 Challenges: Privacy, Policy, and Cultural Responsibility	59
4.3.1 Privacy, GDPR Compliance, and Ethical Data Collection	59
4.3.2 Responsible and Culturally Sensitive Use of AI.....	61
4.3.3. Cultural Stewardship and Sensitive Digital Heritage	62
4.3.4 International Policy, Repatriation Debates, and Digital Data	62
4.3.5. Cybersecurity and Protection of Digital Heritage.....	63
4.3.6 Cross-Institutional Collaboration and Policy Alignment.....	63
Chapter 5 – Conclusion: Results and Future Directions	64
5.1 Results and Contributions.....	64
5.2 Future Research Directions	66
5.3 Concluding Remarks.....	68
References	70
Appendix A: Comparative Metadata Overview of Cultural Heritage Platforms	76
Appendix B: The British Museum: Digital Infrastructure and Data Framework.....	81
Appendix C: Terminology Index.....	86

List of Figures

Figure 1 Types of Data Managed by Large Museums (source: author)	19
Figure 2 Growth of Digitized Objects in Museums (source: author).....	19
Figure 3 Home page of SearchCulture.gr (National Aggregator of Digital Cultural Content).....	79
Figure 4 Search interface of SearchCulture.gr.....	79
Figure 5 Item detail page for a Greek vase (EKT digital record view).....	79
Figure 6 Home page of The British Museum Collections Online portal.....	80
Figure 7 Search Interface of The British Museum/Collections Online	80
Figure 8 Object record showing metadata fields and image viewer	80
Figure 9 British Museum – Simplified Digital Infrastructure Diagram	82
Figure 10 Example RDF triples linking an object to its creator and findspot.....	83
Figure 11 Schematic triple graph: SPARQL query	84
Figure 12 SPARQL query (the equivalent of SQL for RDF/Semantic Web data).....	85

List of Tables

Table 1 Educational resources, archival documentation (source: [BM-n.d.], [VCSR18])	21
Table 2 Comparison of Metadata Practices Across Cultural Heritage Platforms.....	76
Table 3 Core Metadata Fields Used in SearchCulture.gr.....	77
Table 4 Core Metadata Field used in the British Museum Online Collection.....	78
Table 5 British Museum – Simplified Digital Infrastructure Diagram	81

List of Abbreviations

Abbreviation	Explanation
AI	Artificial Intelligence
API	Application Programming Interface
AR	Augmented Reality
CMS	Content Management System
CRM	Conceptual Reference Model
CSV	Comma-Separated Values
DAM	Digital Asset Management
GDPR	General Data Protection Regulation
GLAM	Galleries, Libraries, Archives and Museums
IoT	Internet of Things
JSON/JSON - LD	JavaScript Object Notation/Linked Data
KG	Knowledge Graph

LOD	Linked Open Data
LLM	Large Language Model
MP3/MP4	Digital Audio / Video File Formats
OBJ/STL/PLY	3D File Formats
SQL	Structured Query Language
TIFF/JPEG	Image File Formats
VR	Virtual Reality

Chapter 1 Introduction

This opening chapter establishes the foundation of the thesis by situating museums within the broader context of the digital age and outlining the scope of the research. It begins with an overview of the data-driven transformation of museums, highlighting how digital infrastructures, big data, and semantic technologies are reshaping cultural heritage practices. The British Museum is then introduced as the central case study, with a justification for its selection based on its scale, digital maturity, and leadership in open-access initiatives. The chapter further defines the research aims and objectives, articulates the research questions guiding, and underscores the ethical and technical considerations that inform the study. To support this framework, figures and tables are included to illustrate the diversity of museum data types, their formats, and their functions. Together, these elements provide the conceptual and methodological groundwork for the analysis that follows in subsequent chapters.

1.1 Background and Scope of the Study

In the digital age, museums have become dynamic data-driven environments. *Museum data* refers to the vast and varied information that cultural institutions collect, manage, and generate through their operations. This includes digitized object records, high-resolution images, metadata, conservation reports, visitor analytics, educational resources, and user interaction logs. As collections are digitized and museums expand their digital presence, the volume, diversity, and complexity of their data has increased dramatically.

The growing reliance on digital systems has shifted museums from being primarily physical spaces of curation to hybrid institutions that manage both tangible and intangible heritage through data infrastructures. Early research highlighted this shift, describing museums as evolving information organizations whose value increasingly depends on how effectively they structure and share data. Since then, digital technologies have further transformed museum practice, enabling interactive exhibitions, personalized learning, virtual experiences, and public participation in curation and access.

Today, the term *big data* has entered the vocabulary of cultural heritage institutions. While traditionally associated with corporate or scientific fields, big data in museums refers to the scale and complexity of digital collections, the speed of data generation (e.g., through web and social analytics), and the challenge of managing and making sense of it all

[Kit14]. Institutions like the British Museum are at the forefront of this evolution, offering open access to millions of records and experimenting with digital platforms to reach global audiences.

Recent scholarship emphasizes that big data in museums is not merely about scale but also about interoperability, ethics, and public engagement. The Museum Big Data 2024 conference highlighted how linked data, knowledge graphs, and AI are increasingly shaping cultural heritage practice, making collections more discoverable and meaningful for diverse stakeholders [PD24]. At the same time, applied computing strategies such as semantic markup and structured data annotations have been proposed as vital for enhancing the online accessibility of museum resources [AGP+24].

More recent frameworks, such as those discussed by Vermeeren, Calvi, and Sabiescu [VCS+18], emphasize the design of digital museum experiences that are user-centered, technologically innovative, and responsive to the needs of diverse communities. This holistic approach to museum data highlights the importance of aligning data management with accessibility, engagement, and institutional mission.

This thesis explores how one of the world's largest museums, the British Museum, manages and utilizes its data ecosystem. The study investigates the types, volume, and strategic use of its data and examines how large-scale cultural institutions adapt to the challenges and opportunities of the digital age.

1.1.1 The Data-Driven Museum Landscape

Modern museums operate in a landscape where data plays a central role across nearly every function. As cultural institutions expand their digital presence, data is no longer a byproduct of museum operations – it is a core resource. Museums now collect, analyze, and interpret vast datasets related to their collections, audiences, educational activities, and digital engagement.

Data is first and foremost essential in collections management. Digitization efforts over the past decades have resulted in millions of object records being documented and made accessible online. These records include metadata, provenance, condition reports, images, conservation history, and interpretive content. Managing this data efficiently requires robust systems such as Collection Management Systems (CMS), Digital Asset Management platforms (DAM), and content delivery networks.

Beyond internal documentation, museums increasingly use data to **understand and engage their audiences**. Tools for tracking visitor behavior – from on-site ticketing systems to web analytics and social media insights – help institutions tailor exhibitions, educational programs, and marketing strategies. This shift is part of a broader movement toward evidence-based museum practice [VCS+18], where decisions are grounded in real user data.

Furthermore, emerging technologies such as augmented reality (AR), artificial intelligence (AI), and personalized recommendation engines are built on data infrastructure. For example, the British Museum's digital strategy includes interactive maps, multimedia guides, and open APIs that allow developers to access their public collection data [BM-n.d.]. Recent discussions at the Museum Big Data 2024 conference stress that the real power of these technologies depends on interoperable data models and ethical governance: linked data, knowledge graphs and semantic modelling are becoming central to making collections both machine-readable and meaningful for diverse audiences [PD24]. Practical methods for increasing findability – such as semantic markup and JSON-LD annotations – have also been proposed as straightforward, high-impact steps museums can take to make their digital collections discoverable by search engines and external services [AGP+24].

As Kitchin [Kit14] argues, the ability to derive knowledge from data is not just about quantity, but about the institutional capacity to store, interpret, and strategically apply that data. Museums that embrace this data-driven model can increase accessibility, foster global participation, and expand their educational impact.

The data landscape of museums is therefore multifaceted – spanning preservation, engagement, interpretation, and innovation. As cultural heritage institutions continue to evolve, their ability to navigate this digital terrain will determine their relevance and resilience in the 21st century.

1.1.2 Why the British Museum? Justification of the Case Study

The British Museum is one of the world's largest and most significant cultural institutions, making it an ideal case study for exploring the management of large-scale museum data. Founded in 1753, it houses over 8 million objects spanning more than two million years of human history, representing cultures from every [BM-n.d.] This vast

collection offers a unique opportunity to examine the challenges and strategies involved in handling extensive and diverse digital data.

In recent years, the British Museum has been a pioneer in adopting digital technologies to increase public access and engagement. Its extensive digitization efforts have made millions of objects available online through an open-access platform, allowing researchers, educators, and the public to explore its collections remotely. The museum's commitment to transparency and accessibility aligns well with current trends in open data policies, making it a relevant and forward-looking example for this study. The British Museum's open-access initiatives and experimentation with platform-based services place it squarely within contemporary practices emphasizing not only large-scale digitization but also semantic enrichment and public-facing discovery services [PD24, AGP+24].

Furthermore, the British Museum regularly reports on its digital initiatives and visitor engagement in its annual reports, which highlight milestones in collection digitization and the implementation of new technologies aimed at enhancing user experience [Tap24]. These efforts position the museum as a leader in integrating technology with cultural heritage preservation.

Analyzing its data ecosystem not only provides insights into the scale and complexity of museum data but also highlights emerging practices in data governance, visitor analytics, and digital engagement that can inform similar institutions worldwide. Overall, the British Museum's scale, digital maturity, and openness make it a compelling subject for investigating how big data is reshaping museum operations and public access in the twenty-first century.

1.1.3 Research Aims and Objectives

The research examines how principal museums process, interpret, and strategically leverage the vast amount of data they produce. Using the British Museum as a case study, this thesis explores the changing role of data within collection management, visitor engagement, and institutional strategy.

The primary objective is to **understand the richness and extent of museum information**, both internal (e.g., collection metadata, conservation records) and external (e.g., visitor statistics, social media activity) sources. Museums increasingly find themselves operating in so-called "data-rich environments" [Gia12], where digital content interactions with

visitors can be monitored, analyzed, and applied in the formulation of curatorial and operational strategies.

The second objective is to **explore how such data is being governed**, with particular reference to the systems, standards, and infrastructures found at the British Museum.

This includes an analysis of open data schemes, the use of APIs and digital preservation practices, and the consequences of data silos, legacy systems, and interoperability.

There is a third objective, to **look at the greater potential of museum data**, especially to improve more accessible, inclusive, and personalized visitor experiences. As digital transformation accelerates, there is growing interest in how technologies such as artificial intelligence, recommendation tools, and virtual exhibitions can innovate and improve the way audiences interact with museums [Par13].

Finally, the thesis aims to contribute to the current discussion about **ethical and responsible data use** in cultural institutions, touching on issues such as digital colonialism, privacy, and cultural sensitivity. To reflect the conference emphasis on practical accessibility solutions and governance, this thesis also aims to **assess how semantic enrichment and structured-data practices (e.g., Schema.org / JSON-LD) can improve discoverability and reuse of museum data and evaluate ethical and legal constraints such as GDPR when operationalizing these practices** [AGP+24, Ped24].

1.1.4 Research Questions

To address this question, the following research questions have been crafted:

- What types of information are produced and managed by the British Museum, and how do they represent the institution's curatorial and operational interests?
- How is this data stored, organized, and made accessible – within the institution and to the public – on digital platforms, systems, and open-access initiatives?
- What are the legal and ethical constraints on the way data from museums can be opened up and reused (e.g., GDPR)?
- How can museum data be effectively utilized in a strategic way to improve visitor engagement, inform educational outreach, and promote institutional transparency?

- How can semantic and linked-data approaches maximize discoverability and reusability in ways that are sensitive to data-protection and ethical considerations?

Each of these questions is not only asking about the technical subjects of data management - e.g., infrastructure, standards - but also about the broader strategic potential which it creates, i.e., public participation and personalization. They also indicate the ethical and legal concerns which have come centrally into most current debate in the entire museum and cultural heritage field [PD24, Ped24].

1.2 Figures and Graphs

Visual representations are essential for illustrating the scale and diversity of museum data. In the context of large institutions like the British Museum, data encompasses multiple domains, including collection metadata, digital images, visitor analytics, and social media engagement. Incorporating figures and graphs enables clearer communication of these dimensions and supports the argument that museum data today qualifies as “big data” in terms of volume, variety, and velocity [Kit14].

One key visual is a **pie chart** demonstrating the distribution of digital resources, such as object records, multimedia files, and curatorial documentation. Another useful figure is a **line graph** showing the exponential increase in digital collection entries over time, reflecting the institution’s digitization efforts [BM23]. Such visuals not only provide context for the magnitude of data managed but also highlight how data creation and access have evolved in response to technological advances and public demand.

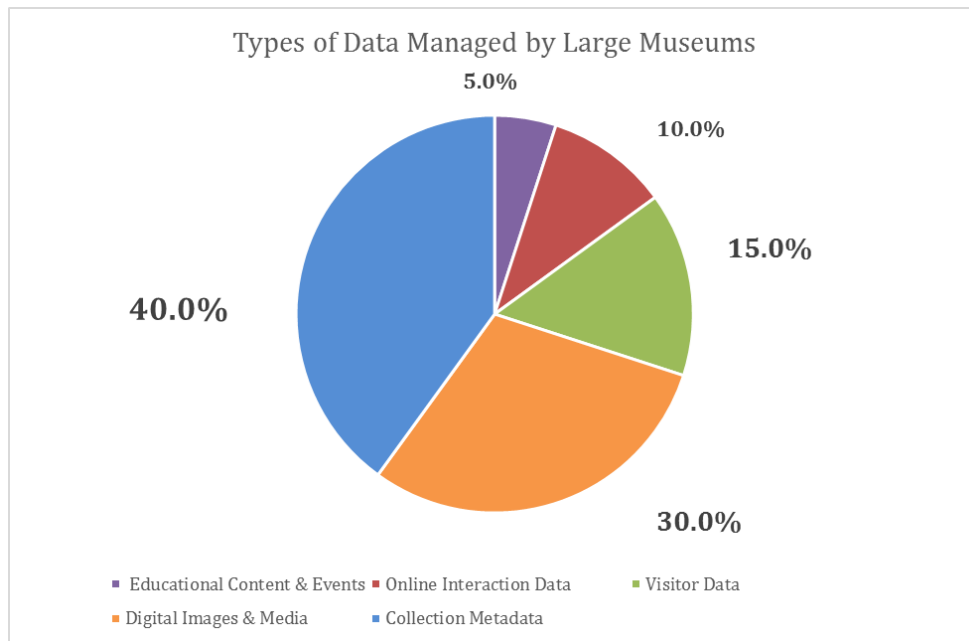


Figure 1 Types of Data Managed by Large Museums (source: author)

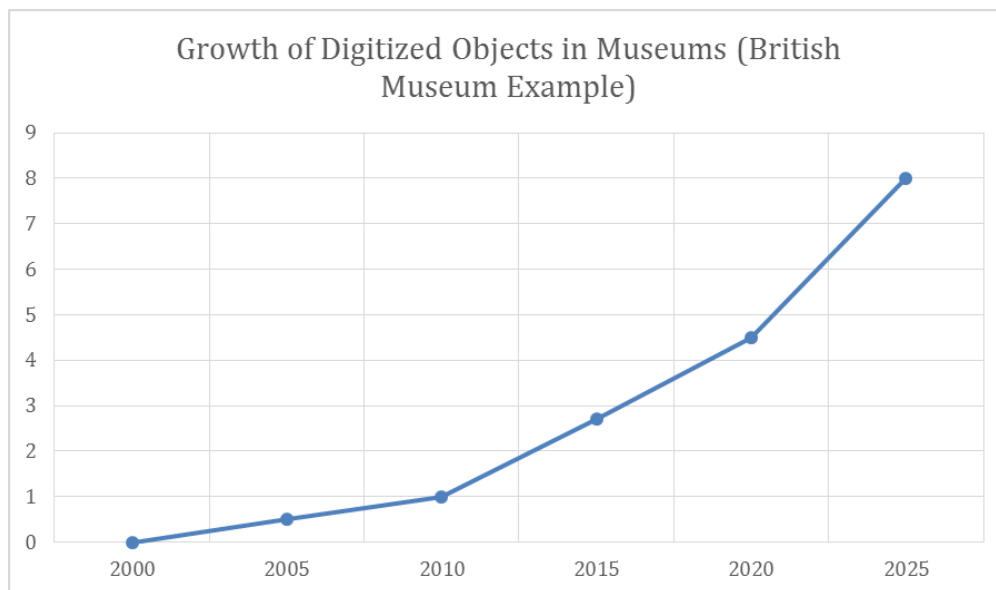


Figure 2 Growth of Digitized Objects in Museums (source: author)

1.3 Tables

Museums like the British Museum manage a wide variety of data, ranging from collection metadata to visitor interactions and digital media. Understanding these data types, their formats, and their primary uses is essential for appreciating both the scale of museum operations and the challenges of managing large digital collections. Tables and structured representations allow for a clearer understanding of the institutional data ecosystem, showing the types of information collected, how it is stored, and what purposes it serves.

The main categories of museum data can be summarized as follows:

Collection Metadata: These are structured records containing information about each object, including title, provenance, materials, dimensions, condition, and historical context. Metadata is typically stored in formats such as XML, JSON, or CSV, which allow for standardized cataloging, easy querying, and interoperability across systems.

Digital Images and Media: High-resolution photographs, illustrations, 3D scans, and videos are associated with collection items. File formats include JPEG and TIFF for images and OBJ, STL, or PLY for 3D models. These resources support online collections, publications, virtual exhibitions, and research activities.

Visitor Analytics: Museums collect data about attendance, ticketing, and demographic information. Databases and CSV files store this data, which is crucial for understanding audience behavior, planning exhibitions, and improving visitor services.

Social media and Web Interaction Data: Online engagement through museum websites, APIs, and social media platforms provides insights into public interest and participation. JSON-formatted data from APIs allows for audience analysis, outreach, and evaluation of digital programs.

Audio and Video Archives: Museums maintain educational and archival media in formats such as MP3 and MP4, supporting both in-person and online learning initiatives.

Data Source	Format	Use Case	Access/ Sensitivity
Metadata collection	XML, JSON, CSV	Object descriptions, provenance, cataloging	Public / Restricted
Digital images	JPEG, TIFF	Online collections, publications, research	Public/ Restricted
3D scans/ models	OBJ, STL, PLY	Virtual exhibitions, preservation, education	Public / Restricted
Visitor analytics	SQL databases, CSV	Attendance tracking, audience insights	Restricted / Internal
Social media data	APIs (JSON)	Engagement analysis, outreach	Public (but privacy-sensitive)
Audio/ video archives	MP3, MP4	Educational resources, archival documentation	Public / Restricted

Table 1 Educational resources, archival documentation (source: [BM-n.d.], [VCSR18])

This table provides a clear framework for understanding the types of data managed, their formats, and how they support both operational and strategic museum functions. It also sets the stage for later chapters, which will analyze data management practices, digital infrastructure and strategic utilization.

Chapter 2 Literature Review

This chapter outlines how museums have evolved into data-rich ecosystems, where digitized collections, visitor analytics, and multimedia assets shape practice and strategy. We first trace the shift to data-driven operations and the opportunities this creates for access, participation, and personalization. We then examine the frameworks that make this sustainable—digital curation, standards, and governance—before addressing the ethical landscape of open access, GDPR, and responsible analytics. Together, these sections set the foundation for understanding how large institutions like the British Museum manage, connect, and ethically mobilize “big” cultural data.

2.1 Museums in the age of Big Data

In recent decades, the idea of “Big Data” has moved beyond business and science and into the cultural heritage sector. Museums are no longer only physical spaces that safeguard artifacts; they have also become complex digital ecosystems that generate, store, and analyze vast amounts of information. This includes digitized collections, visitor analytics, social media interactions, and multimedia content. Early study identified museums as evolving information institutions, and more recent research shows that data now supports nearly every aspect of museum operations [Kit14]. The increasing reliance on large-scale data infrastructures is reshaping how museums engage with audiences, managing their resources, and plan for the future. As noted in the Preface to the *Museum Big Data 2024* conference [PD24], the sector is undergoing rapid transformation, marked by new opportunities and equally significant challenges.

2.1.1 The Shift to Data-Driven Museums.

Museums have progressively moved from traditional physical archives to digital and hybrid models that combine on-site and online experiences. Historically, collections and archives were confined to physical spaces with limited public access, but digitization has helped break down these barriers. A prime example is the Pancyprian Gymnasium Historical Archive in Cyprus, a centuries-old school archive that was once accessible only in person during narrow visiting hours [AKB24]. Many of its documents are fragile and dispersed in storage, hampering research and risking damage from handling. Through a targeted digital transformation, the archive’s contents were digitized and published via the CyprusArk content management system (CMS). This initiative dramatically

expanded access to the archive's rich collections by making them available "anytime, anywhere,online"[AKB24].

The CyprusArk platform consolidated the dispersed items into a single organized repository with improved search functionality, while also protecting delicate originals from physical wear. This case illustrates the broader trend: museums and archives are leveraging digital platforms to preserve heritage and democratize access on a global scale [APK24].

Complementing individual institutional efforts are large cultural data aggregators that epitomize the data-driven museum era. Europeana, the pan-European digital heritage library, and national aggregators like SearchCulture.gr in Greece have transformed how people discover cultural collections. Launched in 2015, SearchCulture.gr serves as the Greek national aggregator and a data provider to Europeana [AGP+24]. Over a decade, it built a robust infrastructure and a public portal that today provides unified access to nearly 1 million digitized objects from 94 different museums, archives, libraries, and galleries. Its success lies in a sophisticated semantic enrichment strategy: SearchCulture.gr applies state-of-the-art linked data techniques to harmonize metadata and enable rich discovery features such as interactive maps and thematic virtual exhibitions [AGP+24].

This shift, as highlighted in the Preface to the MBD2024 conference, is part of a broader recognition that museums must build critical infrastructures and interoperable frameworks to handle the volume, variety, and complexity of their data [PD24]. Museums are increasingly not only physical repositories but also data-driven organizations sustaining large digital ecosystems.

2.1.2 Opportunities of Big Data in Museums

Enhanced accessibility and discoverability. Big Data provides museums with new opportunities to make their collections more accessible to both researchers and the general public. By embedding semantic markup -using ontologies like Schema.org with JSON-LD- into websites and databases, museums can help search engines and cultural aggregators better index and interpret their content [APK24]. This increases visibility in search results, improves interoperability, and amplifies a museum's global reach. Angelaki et al. [AGP+24] demonstrate this practically: integrating Schema.org annotations within a Django web framework significantly improved cultural content discoverability.

Personalized and participatory experiences. Beyond access, Big Data reshapes the museum experience itself. Museums now analyze web traffic, ticketing, mobile app usage, and social media interactions to tailor exhibits and programs. These data-driven approaches make it possible to offer personalized recommendations and create interactive, dynamic exhibitions that evolve with audience engagement. A striking example is the conceptual Folk Tale Museum in Zagori, Greece. In this scenario, big data acts as a “silent co-designer,” with visitor interactions feeding back into the exhibition design to create adaptive and participatory storytelling environments [Pap24]. This model shows how data transforms museums from object-centered repositories into visitor-centered experiential spaces.

The Preface to MBD2024 emphasizes that immersive and participatory technologies, powered by Big Data, are becoming vital for engagement. Data is not only an archival tool but also a creative instrument for shaping narratives and visitor journeys [PD24].

2.1.3 Challenges of Big Data in Museums

Despite all these opportunities, Big Data in museums also creates tremendous legal, ethical, and technical challenges:

Privacy and GDPR: Collecting visitor analytics and personal data raises challenges under moral codes like the EU's GDPR. Museums must address concerns about consent, anonymization, and ownership of visitor-generated data with caution in order to practice compliant and responsible behavior. Peditaki [Ped24] highlights that GDPR compliance often lags the rapid adoption of AI-driven data collection, leaving museums struggling with legal ambiguity [Ped24]. The challenge is not only regulatory but also ethical: museums risk eroding public trust if visitors feel they are being watched too much.

Metadata heterogeneity and interoperability: Museum data is often messy, inconsistent, and non-standardized. SearchCulture.gr illustrates this issue: ingesting records from nearly 100 institutions revealed problems such as inconsistent spelling, duplicate artist names, and vague dating conventions [AGP+24]. To overcome this, the aggregator developed semantic enrichment pipelines that mapped local terms to standardized vocabularies. While effective, such processes demand high technical expertise and resources – a major barrier for smaller museums.

Sustainability and infrastructure: Managing “big” cultural data requires investment in storage, software, and skilled staff. Papaioannou & Damigos [PD24]. Without sufficient resources, museums risk creating large datasets (such as 3D scans or visitor logs) that may become unusable over time due to format obsolescence or the inability to properly maintain them

The Big Data era is transforming museums profoundly. On the positive side, it opens to institutions greater access, more powerful storytelling, and more personalized visitor experiences. On the negative, however, there are concerns about privacy, technical hurdles, and sustainable infrastructure requirements. Museums must find a balance between innovation and responsibility: the promise of Big Data is in its ability to harness it for the purpose of advancing their mission of education and preservation without compromising trust and inclusiveness. These challenges give way to the second section (2.2), where data governance and digital curation will be addressed—explaining how museums can create the infrastructure and policies necessary to effectively manage their growing data environments.

2.2 Digital Curation and Data Governance in Cultural Institutions

This section examines how museums move beyond simple digitization to steward data as a long-lived, reusable asset. We outline the core tasks of digital curation, then show how shared standards—CIDOC-CRM, AAT, Schema.org—enable interoperability and discovery across institutions. We compare platforms and systems (from domain-specific CMS to “digital twin” solutions) that scale documentation, preservation, and access. We also address the rules and responsibilities that frame this work—governance, policy, and legal obligations—before highlighting emerging practices in education and immersive media that activate curated data for learning and public engagement.

2.2.1 Digital Curation in the Museum Ecosystem

Museums now manage far more than physical objects. The rapid expansion of digitization has generated “big data” ecosystems consisting of object metadata, multimedia documentation, visitor statistics, and interpretive content. As Chagas et al. highlight [CFB+24], digitization is not enough; institutions must curate their data so that it is accessible, usable, and useful to education, research, and engagement. This reflects the shift from merely building digital substitutes to overseeing their reuse and long-term worth. Museums have also started to adopt data governance, i.e., regulations and

standards that influence how cultural data is handled, preserved, and disseminated. Vermeeren et al. [VCS+18] detail this in light of a more general transition away from collection-centered to ecosystem-centered museums, as digital infrastructures unite institutions, communities, and external partners. Good governance makes sure this increased participation occurs responsibly, balancing openness with institutional control.

2.2.2 Standards and Frameworks for Digital Curation

Effective curation relies on shared standards that ensure consistency and interoperability. One of the most significant is the CIDOC Conceptual Reference Model (CIDOC-CRM), which models relationships between cultural entities. Nikolaidou [Nik24] applied CIDOC-CRM in documenting artists' studios, showing how it captures both tangible objects and intangible processes, especially when combined with vocabularies like the Getty Art & Architecture Thesaurus (AAT). These ontologies enable integration of data across institutions, supporting sustainability and reusability. In parallel, web standards like Schema.org markup play a growing role in discoverability. The CyprusArk CMS, used for the Pancyprian Gymnasium archive, embeds semantic annotations to make cultural data machine-readable and discoverable through search engines [AKB24]. This not only supports access but situates museum collections in the broader Semantic Web. Together, ontologies and metadata schemas ensure that museum data is not locked in isolated systems but becomes part of a global cultural information infrastructure.

2.2.3 Platforms and Systems for Managing Museum Data

Museums utilize different Content Management Systems (CMS) to host and publish their digital collections. Although generic open-source solutions such as Omeka and DSpace are dominant, they lack domain-specific functionality. The CyprusArk CMS demonstrates how custom solutions can be more suitable for heritage institutions. It features scalable storage and definable schemas, enabling curators to catalog specialist items (coins, manuscripts, etc.) and handle millions of records [AKB24].

Beyond CMS, even more advanced platforms are in development. The ARTES project, presented by Koutoupas et al. [KAB+24], envisions digital twins of art – data-enriched surrogates fusing provenance, conservation, and market information. Equipped with embedded big data analytics and blockchain-based smart contracts, ARTES takes transparency and provenance tracking to a new level, offering governance tools for both curators and the art market.

These examples reveal a spectrum: from practical open-source CMS for small archives to experimental platforms using AI, IoT, and blockchain for managing large-scale artworks. Both approaches highlight how digital infrastructures are central to museum curation in the big data era.

2.2.4 Data Governance and Policy in Cultural Heritage

Governance structures determine who has access to museum data, for what it is employed, and under what conditions. Institutionally, there are permission hierarchies implicated in governance: for example, within the CyprusArk system, teachers and curators each had their own user roles to ensure accuracy and accountability in entering data [AKB24].

Externally, governance is influenced by legal frameworks. Chatzidi identifies the pressures placed on museums by global conventions and domestic legislation fighting against illicit trade in artifacts [Cha24]. Since the 1970 UNESCO Convention, museums have had to record provenance with unprecedented depth, generating large datasets that must be controlled carefully.

Governance also pertains to open access policies, weighing public benefit against ethical concerns such as cultural sensitivity or privacy. As Vermeeren et al. [VCS+18] note, museums must operate within the realm of openness and curatorial authority, deciding what can be opened and what should be closed. In general, governance ensures that museum data are not only technically handled but also legally and ethically proper.

2.2.5 Emerging Approaches: Educational and Immersive Curation

Recent projects show how digital curation and governance extend into education and immersive technologies. In Brazil, the Goiás Memory Center project created a Local History Teaching Repository that reuses museum data for schools [CFG24]. This model integrates academics, teachers, and students into a governance ecosystem, ensuring curated content meets local educational needs. By treating teachers as co-curators, the project shows how governance can be participatory and socially impactful.

Meanwhile, immersive approaches such as VR reconstruction provide new ways of curating archaeological data. Artopoulos et al. [ALD24] describe a collaborative VR environment for reconstructing the Neolithic site of Khirokitia. The platform served as both a research tool (for archaeologists testing hypotheses in real-time) and a public

experience (a virtual tour of the site). This dual function illustrates how digital curation now includes versioning, transparency, and accessibility for immersive data objects.

Both instances highlight that today's curation is not just about storing it is about bringing data to life in a manner that enables learning, collaboration, and public participation.

Digital curation and data governance now play a vital role in the daily operations of museums. Standards such as CIDOC-CRM and Schema.org enable institutions to share and interpret data across diverse systems, while platforms like CyprusArk and ARTES provide adaptable infrastructures for managing and preserving collections. At the same time, governance frameworks help museums navigate their ethical and legal responsibilities. Emerging educational and immersive technologies further demonstrate the value of well-curated data, revealing its potential to enrich learning, support storytelling, and foster more meaningful engagement with audiences.

2.3 Open Access, Data Ethics and Visitor Analytics

This chapter explores how museums navigate the intersection of open access, ethics, and visitor-centric data. As greater collections are going online through cultural institutions, they have new responsibilities regarding privacy, consent, and community rights. At the same time, technological advances in analytics and AI provide opportunities for personalized experiences, while open cultural data allow learning and wider participation. These benefits must be weighed against ethical and legal safeguards to ensure digital transformation builds trust, openness, and the public mission of the museum.

2.3.1 Open Access and Ethical Responsibility

Museums today have a dual mandate of opening access to their digital collections and upholding ethical responsibilities to content creators, communities, and subjects. Opening museum data and collections to open reuse – e.g., through publishing high-resolution images or datasets under open licenses – can democratize the cultural heritage by facilitating public reuse, creative commons, and reuse consistent with the museum's educational mission at museum curatorial research. Increased openness allows audiences not only to consume cultural content but also to remix and interpret it, aligning with the evolving idea of the museum “visitor” as an active participant in the new digital cultural. However, the push for openness brings ethical dilemmas. Curators must balance the public's right to access heritage against the rights of individuals and source communities,

especially when dealing with sensitive material or personal data. There is often a conflict between legal mandates and ethical imperatives: for example, *General Data Protection Regulation* (GDPR) requirements protecting personal data may clash with open-access and reuse initiatives involving community-contributed. Questions arise over who gets to decide the “public interest” in making data openly available versus who has sovereignty over the digital representations of their culture and personal. In practice, museums are learning to navigate these contradictions through *negotiated ethics*, seeking consensus and balancing stakeholder needs. Ethical digitization frameworks stress transparency, community consultation, and sensitivity to context, ensuring that increased access does not become exploitation. Theoretical perspectives underline these tensions. Museum work with digital media as contradictory to participation and representation and asks what genuine participation is like in this new mediascape. Thus, museums must strive for a balance between open access and ethical responsibility, so that cultural data democratization is not done at the expense of privacy, cultural rights, or trust. The balance must be established with clear policies and value frameworks so that open data projects proceed on an ethical, inclusive manner.

2.3.2 GDPR, Privacy, and Consent in Digital Collections

Personal data governance in digital museum projects has been a critical concern in the big data era. The European GDPR and similar laws around the world impose high levels of privacy, consent, and data protection requirements directly impacting how museums collect, store, and utilize visitor information and user-generated content. Personal data may be website and app usage data, usage of interactive exhibits data, or even biometric data for immersive experience in museums. Pediaditaki [Ped24] highlights that the interaction of data protection law and cultural heritage is complicated, especially after the accelerated digital engagement during the COVID-19 pandemic. Her analysis suggests that museum professionals must apply GDPR principles together with developing frameworks (such as the draft EU AI Act) in their handling of visitor information, though these legal regimes "are not necessarily serving the same purpose.". Its core challenge is balancing the individual's private interest (their right to personal data control and privacy) with the public interest of cultural organizations to utilize big data for interpretation and outreach. For instance, museums gather enormous quantities of data from apps and in-museum sensors to study visitor behavior, but under GDPR must make sure such data is anonymized or collected with explicit consent. In fact, questions of “data ownership” have emerged: can data gathered by museum apps be “owned” by someone, or is it a public resource (*res nullius*)? Pediaditaki argues that clarifying who

holds rights to this data – the user, the museum, or third-party providers – is fundamental to lawful and ethical data management. GDPR demands that museums implement data minimization, have consent for processing personal information, and allow people access or erasure of their data. These are sometimes restrictive of digital projects (e.g., limiting the granularity of visitor tracking), but they are necessary to make the public trust them. The MBD2024 conference findings reinforce that ethical reasoning around data must be integral to design, not an afterthought. The conference highlighted the importance of legal and ethical literacy training for museum practitioners in order to use privacy legislation confidently and think creatively about data. Last but not least, GDPR compliance and good consent practice are not just legal requirements but part of museums' ethical responsibility to their publics. By integrating privacy considerations into the early stages of digital collecting, museums can harness the benefits of big data and analytics while upholding citizen rights and promoting a culture of transparency and accountability.

2.3.3 Analytics, AI, and Visitor Personalization

Museums are finding new ways to use visitor data and emerging technologies to make their spaces more engaging, relevant and with optimized operations. Big Data tools are now helping museums learn more about what their audiences are interested in. By looking closely at patterns from things like ticket sales, exhibition interactions, or even online feedback, museums can make better decisions about what to offer and how to present it. One notable trend is the growing use of personalised tools powered by AI, designed to adapt to individual visitor interests – something we've already experienced as very successful in the travel industry. Deliyannis et al. [DMP+24] describe a tourism recommendation platform (INDIANA) that leverages real-time data and AI to offer travelers highly personalized suggestions for cultural points of interest. Central to this system is the creation of a "Digital Twin" – an anonymized digital profile of the visitor that aggregates demographics and behavioral data (via smartphones, wearables, etc.) to fuel recommendations. This concept demonstrates how personalization can be achieved *while preserving privacy*, by using unique identifiers and real-time data without exposing the individual's identity. In a museum setting, such AI-powered personalization might imply suggesting exhibitions, narratives, or educational materials to match every visitor's interest and history with the museum. In fact, the MBD2024 conference identified visitor personalization as a central area of focus for museum AI applications, offering to "increase search, classification, storytelling, and visitor personalization" with strong frameworks in place to govern their deployment. The benefits of this kind of approach

are obvious: the visitor enjoys more interesting and relevant experiences, such as tailored routes through exhibits or augmented reality content that is in tune with his profile and immediate context (e.g. location, time of day). Moreover, museums and local cultural organizations also gain from meaningful insights derived from an aggregated set of visitors' data, to further build up their services and to promote sustainable tourism by dispersing visitors away from popular spots. However, analytics and AI use also raises significant ethical and practical concerns. One concern is privacy and consent in data collection for personalization – anonymized profiles must still be handled in line with data protection guidelines to ensure user trust is maintained. Another factor to consider is how to handle biases in AI algorithms: If the data or algorithms reflect cultural bias or missing perspectives, the system could end up reinforcing stereotypes or leaving out important stories—something museums with a strong ethical focus should work to prevent. Furthermore, there is a disconnect in reality between where sophisticated AI is headed and what the majority of museums can actually accomplish; small institutions such as the Centre of Visual Arts and Research (CVAR) do not often have the resources or expertise to make effective use of analytics, resorting instead to simple tools like social media metrics. Despite this, CVAR's case illustrates the power of analytics: by analyzing Facebook Insights data (e.g. post engagement and follower demographics), the museum could measure levels of audience engagement and refine its digital marketing practices. As museums increasingly adopt such data-informed practices, the field is centering on best practices: integrating ethical reasoning into AI tool design, ensuring transparency in how visitor data is used, and keeping a human-centered approach so that personalization augments rather than dictates the museum experience. In summary, analytics and AI-driven personalization hold great promise for enriching visitor engagement in museums, but their deployment must be accompanied by diligent attention to privacy, fairness, and alignment with the museum's educational mission.

2.3.4 Open Cultural Data: Education, Participation and Interpretation

Open cultural data initiatives in museums are transforming how people learn about and connect with heritage. By sharing their digital collections and metadata for reuse, museums can reach far beyond their own walls, inviting new audiences to explore and engage with culture in creative and meaningful ways. A good example comes from the Goiás Memory Center in Brazil, which shows how open access to museum information can spark interest and learning around heritage. The platform gathers historical collections from local museums and archives and makes them easily available to teachers, students, and community members who once struggled to find or access these

resources. Chagas et al. report that centralizing and openly providing digital heritage materials – complete with thematic indexes and educational guides – helps democratize access to local history, overcoming geographic and socioeconomic barriers to participation [CFG24]. Such efforts align with the principle that heritage education thrives when communities can easily find and use cultural data; as Fratini observes, democratizing access to memory institutions allows citizens (not just researchers) to recognize themselves as stakeholders in cultural heritage.

Open data also leaves space for participatory culture, where users become co-producers of meaning. Students, hobbyists, and teachers are able to remix content into new interpretive products – digital exhibits, research projects, or class materials – whenever museum APIs or open archives are provided, pushing the interpretive narratives away from the museum's curation. This participatory shift is consistent in the fact that the digital mediascape reconfigures the museum's role, refashioning ideas of authority and challenging questions of whose interpretations are important. For example, user-generated content on museum platforms and social media can reflect personal connections to artifacts, adding multi-vocal interpretation layers to the official catalog. In the realm of digital literature, Chalkia et al. [CDS24] note that interactive and data-driven formats blur the line between author and reader, requiring new frameworks for interpretation in literary studies. By analogy, when museums offer open collections, they effectively invite the “reader” (public) to partake in authorship of cultural narratives, necessitating a more fluid, dialogic approach to interpretation.

Moreover, the existence of large cultural datasets allows for data-driven heritage study and new interpretive knowledge. Techniques used in digital humanities like text mining, network analysis, or sentiment analysis can be used for museum data to reveal patterns and stories not visible at the level of the individual object. For instance, combining and comparing metadata of thousands of artworks might uncover trends in iconography across the centuries, or the identification of ignored artists, facilitating more diverse interpretation. Openness to data thus powers scholarly reinterpretation and more substantial storytelling for critical rethinking of digital cultural heritage practice. As cultural data proliferates, museums must develop new theories and curatorial practices to handle “the profusion of new types of societal data” and to remain relevant in a data-intensive society. This includes acknowledging non-traditional forms of heritage (e.g. digital-born art, AI artifacts) and understanding that digital data carries its own context and biases. The Museum Big Data 2024 preface [PD24] also imposes specific emphasis on the point that the growth of museum data (from documentation to social media and

analytics) both offers opportunities and poses challenges in interpretation, governance, and reuse. Overcoming these challenges requires cross-disciplinary frameworks and an ethical, culturally conscious mindset so that open data creates meaningful participation instead of misinterpretation.

In general, open cultural data has the potential to greatly enhance public engagement and education by making heritage resources openly accessible and reusable. It also has the capacity to redirect interpretive practices by spreading the activity of meaning-making across a wider group and enabling new analysis. The compromise for museums is to create these interpretive freedoms and innovations while providing context, direction, and ethical guidance. Doing so, museums fulfill their social role in the digital age: no longer merely collecting and storing things, but curating relations and meanings between people and their shared cultural knowledge. All of these efforts – from open educational environments to co-creative storytelling – indicate how "big" museum data can be ethically employed to build knowledge, cultivate inclusiveness, and facilitate novel meanings for our cultural heritage.

Chapter 3 The British Museum: Data Ecosystem

This chapter maps the museum data ecosystem and anchors it in the British Museum's practice. We first sort museum data into four working groups – collection/object records, visitor and operational signals, digital engagement/media, and research/collaborative outputs – and note typical uses and pitfalls. We then look at the British Museum's digital stack: public portals and APIs, adoption of shared standards and linked data, and trials with VR and AI to enrich access and interpretation. Finally, the chapter compares the British Museum's scale with that of a national aggregator and a smaller institution, showing how an organization's size influences its digital systems, data quality, and ability to reach the public. The goal is to link different types of data and platforms to the everyday work of museums – discovering, preserving, and sharing knowledge.

3.1 Types and Sources of Data

In the field of museum informatics and big data, it's important to understand the different kinds of data that museums create, manage, and use. This section outlines four main types – Collection & Object Data, Visitor & Operational Data, Digital Engagement & Media, and Research & Collaborative Data. Each category of data plays a distinct part in how museums operate and share knowledge. Looking at real cases and examples, we can see how institutions like the British Museum document their collections, observe and interpret visitor behavior, use digital tools to reach wider audiences, and collaborate on research and learning projects. The following sections examine these areas in greater depth, outlining the main approaches, difficulties, and possibilities involved in managing the vast and varied data that museums handle today.

3.1.1 Collection & Object Data

Collection and object data form the foundational layer of museum information. This category covers all information connected to the museum's objects and artworks – from catalog records and provenance details to high-resolution images, 3D scans, and conservation notes. Museums have long documented and organized such data through careful and detailed cataloguing, following shared vocabularies, schemas and data standards to keep records consistent. In recent years, new methods such as semantic frameworks and digital twin technology have further strengthened this process, making object data easier to preserve, connect, and access both within the institution and by the public.

A sample innovation in data gathering is ARTES, a platform that combines big data approaches and digital twin technology in art collection management. The ARTES system was designed to address some of the age-old challenges in art documentation such as authenticity verification and tracking provenance. By creating unique digital replicas, "digital twins", of artworks and linking them to abundant metadata, ARTES improves documentation accuracy and enables real-time analysis of art objects' past and value. As well as making an artwork's provenance and condition more transparent, this approach protects cultural heritage by preserving full information about every object digitally. Basically, ARTES is a digital management pipeline for art ends-to-ends that combines advanced technology and expert analysis to improve collection documentation, confirmation of authenticity, as well as dynamic valuation models. This type of platform illustrates how big data and digital technologies can improve the structure and preservation of museum object data.

Another vital development in this realm is the use of **ontologies and standards** like CIDOC-CRM to model complex relationships around objects and their context. A case in point is the documentation of **artists' studios** as cultural heritage entities. Artists' studios often contain a wealth of tangible and intangible information (e.g. tools, unfinished works, spatial arrangements, personal effects) that traditional catalogs might overlook. Researchers have proposed semantic documentation frameworks to capture

this complexity. For example, Nikolaidou suggests an ontology-informed model for recording artists' studios using the CIDOC Conceptual Reference Model and the Art & Architecture Thesaurus (AAT) [Nik24]. By organizing studio elements and connecting them with the artist's creative process, this approach ensures that ephemeral aspects of an artist's working environment are documented and linked with the museum's collection metadata. The use of CIDOC-CRM in this instance promotes understanding on a shared basis between institutions and supports long-term maintenance of information about the studio within a broader cultural heritage context. In practice, what this means is that not only are the art objects themselves documented, but also the context in which they were created – providing more context and interpretative sense to each piece.

Apart from semantic modeling, museums also make use of Collection Management Systems (CMS) to handle object data in large numbers. For instance, the Cyprus Institute has developed the CyprusArk digital repository (a form of CMS) to gather and archive cultural heritage data in Cyprus. Such systems integrate object records, images, and multimedia, often fusing standards like Dublin Core or CIDOC-CRM in order to facilitate

interoperability. By leveraging a specialized CMS, museums can ensure that the information for each object – ranging from basic catalog records to conservation reports – is securely stored and can be queried or shared as needed. This is especially crucial to a large museum (like the British Museum) that might have millions of object records. The CMS solution complements semantic ontologies: ontologies define data structure and relationships, but the CMS constitutes the platform to store and care for the data from birth to ultimate destruction.

Overall, Collection & Object Data encompasses structured descriptive metadata, provenance and authenticity metadata, and digital manifestations of physical objects. The examples of ARTES and studio documentation using ontology show how embracing big data technologies and semantic standards can bring a significant impact to the documentation and preservation of object-related data by museums. Through this process, the museums not only safeguard the information regarding every artifact but also enhance discoverability and usability of collection data for internal consumption and public access.

3.1.2 Visitor & Operational Data

In addition to collections, museums develop and utilize visitor and operational data – data that describes how people interact with the museum spaces (physical or virtual) and how the institution itself functions. This category can include such things as visitor demographics, attendance, ticketing information, gallery traffic patterns, exhibition popularity, online traffic, and even environmental or IoT sensor data from the museum. Comprehensively understanding this information allows museums to personalize experiences for their public, optimize operations, and guide decision-making in marketing, staffing, or exhibit design.

One of the most important tendencies in this sector is the use of personalization and profiling systems for museum visitors with the aim of enhancing the museum visit experience. One example is the INDIANA project, a pioneer system in using big data and artificial intelligence in museum tours and cultural tourism. INDIANA creates a highly personalized digital guide for each visitor by building a profile often referred to as a “*Digital Twin*” of the traveler. This digital profile aggregates data about the visitor – such as demographic information, interests, location, and even real-time conditions like weather or crowd levels – in order to recommend personalized Points of Interest (e.g. museum exhibits, cultural sites or activities) in real time. Through machine learning and

data analytics, the system can adapt suggestions on the fly, ensuring that the content each visitor sees is engaging and relevant to them. A key innovation of INDIANA is how it integrates multiple data sources (museum databases, tourism information systems, user-generated inputs) while *respecting privacy* by keeping the digital twin profile anonymous. This website benefits not only the visitors but also provides insightful operation information for cultural organizations and museums. For instance, by analyzing the collective visitor information, local museums are in a position to determine trends like which exhibitions are well-liked among certain groups of individuals or how weather influences visiting museums, thereby maximizing their provision and resource use. Also, INDIANA employs IoT sensors (e.g., beacons or smartphone sensors) to collect real-time figures regarding the flow of visitors, enabling features like crowd control – dynamically guiding or limiting visitors to prevent congestion in busy galleries. This is a demonstration of visitor data analytics improving both the visitor's experience (through personalization) and operational efficiency (through real-time observation and reconfigurations) for museums. As the project's early findings show, *“the use of visitor profiling... enables the platform to customize experiences for each visitor,”* while real-time technologies like AR and IoT allow for *“dynamic content adaptation and effective crowd management,”* helping mitigate over-tourism and enhance sustainability of the visitor experience [MFD+24].

On a smaller scale, museums also gather operational data from online engagement and day-to-day activities to inform their strategies. The **Centre of Visual Arts and Research (CVAR)** in Cyprus provides a case study of a small museum leveraging data for operational insights. CVAR, a small organization, focused on website and social media visits to determine the reach and interest of its audience. An analysis of CVAR's online presence (website and Facebook/Twitter sites) revealed that the museum had an adequate informative online presence but lacked a strategic, data-driven strategy to optimize these channels to their fullest potential. Specifically, the report found that CVAR's digital content was largely static and report-like, which, while useful, were not effectively progressing its audience or interacting with the users in a consistent manner. The researchers proposed a more marketing-oriented utilization of the website and social media – like using analytics to identify what kind of content gets people responding, or what virtual exhibitions receive the largest number of hits – and then developing the museum's digital strategy off that. This would allow CVAR to further support its functions such as building audiences, obtaining new funding, and fulfilling its educational mandate. Practically, that means tracking figures like page visits, social

shares, or web inquiries and interpreting them as feedback on what operates well operationally or what the museum can do differently. All these analytics are becoming accessible even to small museums in the form of free or low-cost software (e.g., Google Analytics, social media reporting). The CVAR case points out that data-based decision-making is not only limited to tech giants or national museums, but local museums also gain by fine-tuning their operations in sync with conclusions drawn from visitor data. Lastly, it is to be remembered that managing visitor information has its own responsibility, especially when it concerns privacy and ethics. Museums nowadays need to ensure that any personal information gathered (e.g. via mobile apps, ticketing systems, or Wi-Fi tracking) is processed in accordance with data protection rules. Consent, anonymization, and transparency issues take precedence. For example, personalization systems like INDIANA explicitly design the digital twin to be anonymous, highlighting a privacy-by-design approach. A legal analysis by Pediaditaki [Ped24] also emphasizes that as museums leverage personal data and big data, they need to navigate the complexities of data ownership and GDPR – ensuring that the benefits of visitor data never come at the expense of individual rights.

In summary, **Visitor & Operational Data** is a rich resource for museums to improve experiences and efficiency, but it must be managed strategically and ethically. Real-world projects like INDIANA demonstrate the potential of big data in tailoring cultural experiences, while studies on small museums like CVAR show that even basic analytics can guide better operational decisions.

3.1.3 Digital Engagement & Media

As museums extend into cyberspace, more and more of their information comes from digital engagement and media sources. This comprises all information generated through museums' interactions with citizens by way of social media, mobile applications, interactive experiences, and other open digital systems. It also includes the material created for or by internet users – such as user comments, social media posts about museum exhibitions, digital installations of artwork, and co-storytelling initiatives. Both of these are central in this instance: digital media used by museums to communicate with audiences, and audiences' generation of data and content that may be learned from or utilized by the museum.

Social media data is a direct component of this categorization. Museums use social media platforms like Twitter, Instagram, Facebook, or TikTok in a positive way by sharing

information and connecting with local and global audiences. A click – likes, shares, comments, hashtags – is data that can determine the quality of engagement and what the people think. For instance, more tweets about a new exhibition are real-time feedback for marketers and curators. Most museums also employ social media analytics to learn more about their publics (e.g., viewing which posts create the most buzz or what demographic is engaging with what information). This enables museums to tailor their online outreach and even inform live programming (what's discussed in a chat room might be mirrored or expanded in a physical space). Social media is also a space for participation: fans often append their own photos and stories of trips to the museum, essentially co-authoring the museum's story in the digital sphere. This kind of user-generated content is a great kind of engagement metric that can be monitored by museums and even sometimes reused (with permission) in order to construct a sense of community around the collections [Sim10].

Along with traditional social media, museums are also using interactive and immersive digital experiences to heighten the visitor experience. An unusual example is the museum design concept for a **Folk Tale Museum in Zagori, Greece**, where data-driven interactivity takes center stage. Here, rather than passive recipients of information, visitors become active co-creators who can influence the museum experience in real time. Papadopoulou [Pap24] describes how the proposed Folk Tale Museum would use art and technology to create an *“interactive and inviting environment”* where visitors' inputs (their choices, preferences, and feedback gathered as data) affect both the narrative flow of folk tales and even the spatial exhibition design. In other words, big data is an "invisible co-designer" of the museum experience, changing and adjusting the exhibit material continuously in response to visitor engagement. This participatory approach not only preserves and reinvigorates intangible heritage (in this case, the rich stock of Zagori folk stories) by giving it a life of its own, but also establishes an amplifier effect: visitor preference and behavior are constantly fed into new knowledge that optimizes the museum's provision. The result is a "participatory, adaptive space, where visitors contribute to the very nature of its being.". This kind of design fits into the greater movement of the participatory museum, wherein the audience's participation is at the center of the process of content creation and storytelling. It understands that audiences today, especially young generations, prefer interactive and customized experiences. Leaning on the use of digital media (touchscreens, AR/VR, interactive storytelling software) and having the capacity to read the data that these interactions produce, museums can regularly change exhibitions to remain up to date and engaging.

Digital engagement also applies to the way museums offer narrative content online. **Digital storytelling and literary production** in the museum context can refer to things like online exhibitions, virtual tours with storytelling elements, or even museum-sponsored interactive fiction. A case from the digital humanities – while not a museum exhibit per se – highlights the changing nature of narrative engagement: the rise of **digital literature**. Chalkia et al. (2024) describe how books offered on online media (e-books, hypertext fiction, interactive storytelling, etc.) alter the reader's role from passive to active contributor or co-author. Now the readers are able to choose non-linear paths or even assist in shaping the story, thus demarcating to a certain extent the difference between reader and author. This mirrors what is happening in cultural institutions: so, for example, a museum might have a crowdsourced storytelling project where individuals submit memories or interpretations to accompany objects in an online gallery. The **interpretive skills** of audiences are engaged in new ways – instead of merely consuming a curator's label text, digital platforms invite users to comment, tag, remix, or augment the stories around museum objects. The Folk Tale Museum scenario is one such museological application of this concept, effectively treating folk narratives as an open-ended story that museum-goers help to tell. The "exchange between data and storying has the power to transcend time and place and bring individuals to folklore not as passive recipients but as active co-creators in the process." This quote highlights the magic of digital experience: it is able to turn a visit to a museum or an online cultural visit into a co-creative, interactive act.

In brief, Digital Engagement & Media Museum data involves metrics and content that come from interactive and social technology. Museums translate this data to gauge public engagement and adjust their strategy, and use digital media to promote more engagement. By seeing how visitors engage with museum websites, apps, social media, or pilot projects, museums gain insight into visitors' online interests and behaviors. As the Zagori experiment and other digital literature trends illustrate, taking on participatory design and interactive media can make culture accessible to all and more relevant, enabling a two-way dialogue between museums and the public.

3.1.4 Research & Collaborative Data

Museums do not only collect data for their own use; increasingly, museum data is being shared, reused, and enriched in **research, educational, and collaborative contexts**. This category includes data that museums provide to scholars, the integration of museum datasets with external knowledge systems, and community-driven projects where the

public or other institutions contribute to or utilize museum data. Such collaborative endeavors expand the impact of museum collections, turning isolated silos of information into part of a larger, interconnected knowledge infrastructure. They also illustrate the “prospective” use of museum data – how it can fuel future discoveries, learning, and innovation.

A cutting-edge example in this space is the use of museum data in combination with **Artificial Intelligence and Knowledge Graphs** to create advanced knowledge services. The **GPToLODS+ prototype** showcases how cultural heritage data can be reused in an AI-enhanced context [MKT24]. GPToLODS+ connects a Large Language Model (ChatGPT) with a vast aggregation of Linked Open Data (LOD) sources, including hundreds of RDF knowledge graphs relevant to cultural heritage. Many of these graphs are based on museum and archival data models like CIDOC-CRM, meaning that museum collection data (objects, people, places, events) published as linked data becomes fodder for the AI’s knowledge base. The system enables a range of functionalities useful for both researchers and casual users: for example, one can ask natural language questions and the AI will draw upon museum data from those knowledge graphs to answer, complete with provenance (citations) from the original datasets. It can also perform entity recognition and linking on text (identifying mentions of artists or artifacts in a ChatGPT response and linking them to museum database entries), as well as *fact-checking* ChatGPT’s statements against verified museum records. In essence, GPToLODS+ is a collaborative bridge between the unstructured knowledge of AI and the structured, vetted knowledge in museum databases. By integrating these, the tool enhances discoverability and validation: a user might ask a complex question that requires combining data from the British Museum’s collection, a historical archive, and a library – GPToLODS+ can potentially traverse all those linked datasets to provide an answer, while citing the museum data sources. This example highlights how museum data, when shared in interoperable formats, can be **reused in broader knowledge networks** and even plugged into AI services. It also benefits museums by driving external interest back to their collections (through citations and links) and by helping identify gaps or inconsistencies in their data via AI-assisted analysis.

Another domain of collaborative museum data use is in **education and community history projects**. Museums often collaborate with schools, universities, and local communities to reuse digital collection data for educational purposes. A notable case is the creation of the **Goiás Memory Center** in Brazil, which is part of a local heritage education initiative. In small towns like Morrinhos (Goiás state, Brazil), physical access

to national museums or archives is limited, so digital collections become crucial for teaching history. However, simply having digitized items online is not always sufficient – educators found that many online museum resources were hard to navigate or not aligned with school curricula. In response, Chagas et al. proposed a model for “*heritage digital curation*” that adapts museum data for local educational use. This involves the development of digital items from museum collections (e.g., photographs, texts, regionally associated artifacts) and bundling them into thematic sets and lesson plans that can be readily adopted by teachers and used in the classroom [CFG24]. The project is collaborative in nature: it entails uniting the university (staff and students as digital curators), the museum sector (providing the collection data), and school teachers (who provide feedback on pedagogical needs). The Local History Teaching Repository developed at Goiás Memory Center illustrates the way in which museum information can be reworked and augmented with additional context to meet community requirements. For example, a photo from a national archive could be supplemented with a local oral history recording or map, then displayed as part of an interactive timeline for school students. By making reuse a priority over access per se, this project signals a shift of thinking: museum digital collections are not just end-products to be consumed, but raw materials for the production of knowledge anew. Early results indicate that this kind of educational reuse can make heritage more accessible and relevant to people’s daily lives, reinforcing cultural identity and critical historical thinking in young learners. It’s an excellent illustration of collaborative data: multiple stakeholders (museums, educators, community members) co-create a resource that none could have produced alone.

Community-driven heritage data projects also fall under this category. One example is the data initiative around the martyr village of Kommemo in Greece, a community heavily affected by a historic tragedy (WWII massacre) that has been the subject of cultural memory preservation efforts. Andrianou explored a “*data-driven interpretation of cultural landscapes*” in Kommemo [And24]. While details of that study are beyond this summary, the essence is that community heritage sites generate various data (memorial records, personal testimonies, geospatial data of monuments, etc.) that can be integrated and analyzed to tell a richer story of the place. In such collaborative endeavors, communities contribute data (e.g., families contributing letters or photos, volunteers recording memorial sites) that is supplemented with archival or museum records. The result is a richer documentation of heritage for the use of both scholarship and the needs of the community to remember and interpret their past. The Kommemo case shows the way that

community heritage information can be part of the museum data environment when such information is tackled as a collaboration: local stakeholders and scholars together ensure the validity and pertinence of data, while digital tools (like GIS mapping, data visualization) make information engaging and understandable. Such efforts align with the ethos of open-access and inclusive digital transformation in heritage, as noted by international working groups on Digital Community Heritage.

In summary, **Research & Collaborative Data** covers the outward-facing life of museum data – how it is used in research beyond the museum’s walls, in educational settings, or in community projects. Key to this is interoperability and openness: museums that adopt linked data standards, open licenses, or collaborative platforms find their data more easily integrated into new works. Whether it’s an AI system like GPToLODS+ validating facts against a museum database, or a university partnering with a local museum to create a curriculum, the trend is toward breaking down data silos. This multiplies the value of museum data, enabling new insights (researchers might discover patterns across collections), new learning outcomes (students engaging with primary sources), and new forms of cultural production (communities building their own heritage narratives). Ultimately, treating museum data as a **shared resource** encourages innovation while also fulfilling museums’ mission to disseminate knowledge. As the cases above illustrate, the collaborative use of museum data can lead to sustainable models of cultural heritage preservation and education, ensuring that the data generated by a “big” museum has a big impact on society.

3.2 Digital Infrastructure and Platforms at the British Museum

The British Museum's digital infrastructure facilitates access to its large collections and interaction with global audiences. As one of the largest museums globally, it has developed comprehensive online platforms, used data standards and APIs to facilitate interoperability, and experimented with immersive technologies and AI systems to enhance curation and learning. This section addresses the Museum's digital resources (e.g., web portals and APIs), its use of data standards and linked data, and innovative tools like virtual reality (VR) and artificial intelligence (AI) in its functions.

3.2.1 Online Collection Platforms and Data Access

The centerpiece of the Museum’s digital platform is the **Collection Online** database, a public web portal that allows users worldwide to search and explore millions of objects.

Following a major revamp in 2020, the British Museum's online database provides access to over 4 million object records, a subset of the Museum's approximately 8 million total objects. In practice, this amounts to nearly five million distinct objects present in more than two million digital records, with many of those with high-quality photos and detailed descriptions. Keyword searching as well as filtering by object type, artist, location of production, and date are all available on the site, making both casual browsers and scholars able to look through the extent of the collection. This extensive web presence is an example of how a wonderful museum leverages digital infrastructure to introduce holdings beyond the boundaries of the galleries.

In addition to its main website, the British Museum also extends itself into the digital world through external websites and open data sources. For instance, the Museum makes over 7,000 objects available on Google Arts & Culture, a web aggregator that hosts museum content from around the world. Initiatives like the interactive "Museum of the World" (developed in partnership with Google) uses British Museum data to construct immersive timelines of human history, connecting objects across time and place in a virtual interface.

Such initiatives illustrate the Museum's strategy of meeting users on multiple digital platforms. Furthermore, the British Museum provides programmatic access to its data via APIs and Linked Open Data services. A SPARQL endpoint releases the collection as a machine-readable knowledge graph (semantic web standards-based), and downloadable data dumps of collection metadata for reuse are provided by the Museum. The services enable researchers and developers to query and analyze the complete collection dataset, evidencing open access and interoperability commitment. By maintaining user-friendly web interfaces and robust APIs, the British Museum places its digital base within reach of diverse publics—from the general public to researchers collecting data at scale.

3.2.2 Data Standards and Interoperability

Given the Museum's enormous and heterogeneous collections, adopting common **data standards** and linked data practices is crucial for interoperability. A SPARQL endpoint exposes the collection in a machine-readable form. The British Museum's digital systems follow widely used cultural heritage schemas and ontologies in order to enable its data to be disseminated and understood across broader networks. Accordingly, the database for the collection has been associated with semantic ontologies (e.g., the CIDOC

Conceptual Reference Model) and other Linked Open Data standards so that it can be interlinked into global knowledge graphs of heritage. This is one of a broader trend in the museum world: museums are starting to implement Linked Data and structured metadata to improve documentation and connect records to outside authority files and vocabularies. Papaioannou & Damigos [PD24] note that novel museum projects employ knowledge graphs and ontologies to enhance documentation and make relationships between objects, individuals, locations, and time periods in collections more accessible. In line with these trends, the British Museum has become part of semantic data programs that operate with its collection as part of an ecosystem similar to the web rather than a separate database.

Interoperability also entails adherence to international standards for data exchange. Several aggregators and museums implement standards like the Europeana Data Model (EDM) or Dublin Core for record sharing. The Greek national aggregator SearchCulture.gr, for instance, requires contributing institutions to supply metadata in EDM or equivalent formats by means of protocols like OAI-PMH. This approach streamlines aggregation into European platforms and promotes a common data structure. The British Museum similarly adheres to standard cataloging and licensing practices so that its data can be harvested or linked by external services. In fact, the Museum's collection metadata has been made available as a Linked Open Data dataset, which third parties can query or download under an open license. By doing so, the Museum enables integration with projects like Europeana, and it allows developers to build applications using BM's data without starting from scratch.

A key challenge that comes with big, aggregated datasets is **semantic consistency** – different collections may use varying terms or formats for similar concepts. Advanced digital infrastructure tackles this via **semantic enrichment**. SearchCulture.gr provides a compelling comparison: it developed a “state-of-the-art semi-automatic semantic enrichment” pipeline to homogenize metadata from 94 different providers, covering nearly 1 million objects. This enrichment involved mapping disparate terms to common Linked Data vocabularies (e.g. linking person names to VIAF authority records, or object types to the Getty AAT). As a result, each item in that aggregator is augmented with standardized fields for “what, who, where, when” – improving search precision across the platform. The British Museum's own data curation follows a similar philosophy: by using controlled vocabularies and aligning with external identifiers, the Museum ensures that a search for “Rameses II” or “Amarna” links all relevant objects together even if

originally catalogued under different spellings or languages. In essence, robust data standards and semantic practices.

3.2.3 Immersive Technologies and Virtual Reality

In recent years, the British Museum and other top institutions have incorporated immersive technologies such as virtual reality (VR) and augmented reality (AR) into the museum experience. Top museums with international reputations have introduced VR-based exhibitions and virtual tours, often in the context of temporary exhibitions or special events. The goal is to create rich, interactive experiences beyond those achievable through conventional displays. Similarly, at the British Museum, the Samsung Digital Discovery Centre has acted as an experimentation zone for piloting VR and digital interactives. Worth mentioning, in 2015 the Museum launched a VR experience that transported visitors to a Bronze Age roundhouse and placed 3D-scanned artefacts in their prehistoric context. Visitors may wear VR headsets or step inside an immersive dome to stroll through a realistic ancient dwelling with a smoky fire, observing how artifacts from the collection (the Bronze Age Beaune Dirk dagger, for example) would have appeared in their home context. This innovative event demonstrated the potential of VR to set museum objects in reconstructed environments and expand visitors' understanding by simulation and story.

Application of VR and related technologies represents a broader trend in digital infrastructure to create immersive learning environments. Museums have been using videos and touchscreens for years; VR now offers a more interactive, first-person experience of history. Such reconstructions were found to improve users' cognitive processing of complex spatial and historical information, according to research. For example, nearly replicating an archaeological site or ancient building allows the public and specialists to experiment with hypotheses about how people lived, or how some artifact was used, in ways that cannot be attempted using a fixed exhibit. Artopoulos et al. describe how immersive VR reconstructions were employed in archaeology to collaboratively model a Neolithic settlement, blending research and public education goals [ALD24]. In the museum context, similar techniques are used to create digital exhibits – from virtual reality tours of ancient sites to augmented reality apps that let visitors visualize missing colors on a faded sculpture. The British Museum's infrastructure has evolved to support such projects: it involves not only the VR headsets and software but also the underlying 3D data capture (e.g. photogrammetry and laser scanning of objects) and content management to integrate these immersive experiences

with the collection database. By investing in immersive technologies, the Museum adds a new dimension to its digital offerings, making history accessible in ways that are interactive and emotionally resonant.

3.2.4 Artificial Intelligence and Data-Driven Systems

Some of this digital transformation is also the British Museum's discovery of artificial intelligence (AI) and data analysis to assert control and comprehend its collections. The application of AI in museums ranges from the automation of routine tasks (such as image recognition for cataloging or visitor chatbot tours) to making sense of "big data" collected about objects and visitors. According to recent research, museums are increasingly dependent on techniques like data mining, machine learning, and predictive modeling to uncover hidden patterns in their gigantic datasets. Within the extensive corpus of the British Museum, AI is employed to recognize relationships that wouldn't be obvious otherwise—e.g., clustering artifacts by material or style through image analysis or recommending comparable objects to web viewers based on how they view. Papaioannou & Damigos [PD24] point out that in good integration, AI and big data enable museums to provide richer and more personalized experiences, for instance by dynamically curating content that resonates with the interest of a user. They also point out, however, that such technologies introduce new challenges in respect to data quality, bias, and ethical stewardship in museums. This is a significant factor for the British Museum, which must balance artistic license with its position as a revered institution of culture. Any AI solution applied to interpretation or public engagement is extensively tested to avoid misrepresentation of objects or cultural bias in recommendations from algorithms.

One area where AI-like approaches have already proven valuable is in **enhancing metadata and search capabilities**. The British Museum's collection data, accumulated over 250 years, inevitably contains inconsistencies and gaps that modern algorithms can help reconcile. For instance, the Museum could employ natural language processing to standardize historical descriptions or to link object records with external datasets. The practice of semantic enrichment seen in aggregators is relevant here: SearchCulture.gr used semi-automatic tools to disambiguate names, places, and dates across different collections. That process, essentially an AI-assisted metadata cleanup, significantly improved search precision and user discovery on a platform handling nearly one million items. Similarly, the British Museum can apply machine learning to assist with cataloguing (e.g., object recognition in images or auto-suggesting taxonomy terms) and to emphasize discoveries of potential errors in records. Already demonstrated by

computer vision applied to museum collections to identify artworks or identify features, these approaches could be extended to the BM's extensive collections. Moreover, machine learning-based analysis of user interaction (online and in-gallery) can be employed to identify how the Museum organizes information. For example, grouping user queries can reveal emergent interests which may inform digital exhibition themes.

Central to this, the Museum sees AI as a way to assist human curation, not replace it. Ethical concerns are served: privacy concerns (in the segmentation of user information), intellectual property, and accessibility are addressed by governance policy. The Museum's digital strategy acknowledges that algorithms are programmed by their training data, in the cultural heritage context possibly translating to over-representation of those well-documented items and under-representation of items from underrepresented cultures. Therefore, all AI systems are developed in consultation with curators and data specialists to ensure respect for the range of the collection and for the public purposes of the Museum. Overall, the British Museum's use of AI and data systems is cautious but forward-looking – attempting to fulfill the potential of big data for scholarship and public access without sacrificing demanding ethical standards and scholarly norms in its digital foundation.

3.3 Scale of the British Museum's Collection Data (Comparative Perspective)

The British Museum's digital infrastructure operates at an extraordinary scale, especially when compared to smaller institutions. The Museum's collection comprises around 8 million objects in total, of which roughly half have been catalogued in the online database – approximately 4 million object records accessible to the public. This scale places the British Museum among the largest collection databases globally, both in physical holdings and in digitized content. Managing such volume poses unique challenges: from ensuring data consistency across millions of entries to delivering fast search performance for a worldwide user base. It also offers opportunities – for instance, big data analytics become possible when you have millions of data points, allowing the Museum to observe broad trends in collection usage or to train AI models on a rich dataset. The Museum's ongoing digitization efforts seek to close the gap between the total objects and online records, aiming to eventually bring more of the ~8 million items to the public digital domain.

To put the British Museum's data scale in context, it helps to compare with other platforms and institutions:

- **British Museum (UK):** ~8,000,000 objects in the collection, ~4,000,000 of them with digital records online.
- **SearchCulture.gr (Greece's National Aggregator):** ~1,000,000 cultural heritage objects aggregated from 94 local collections.
- **CVAR Museum (Cyprus, small art museum):** "Thousands" of objects in its collection (on the order of 10^3 rather than millions).

This comparison highlights a stark gradient in data scale. The British Museum's collection size is an order of magnitude larger than a national aggregator portal that combines content from nearly a hundred institutions, and several orders larger than a single small museum. For example, SearchCulture.gr, as a centralized repository for Greek cultural heritage, has amassed about 860,000 records over its first decade. That is an impressive number, yet it is still only about one-fifth of the records available on the British Museum's site. Meanwhile, a museum like the Centre of Visual Arts and Research (CVAR) in Cyprus has a collection modestly described as "numbering thousands" of objects. Such smaller museums do not necessarily have the resources to digitize and make available large databases, and their online presence is less extensive. In the CVAR case study, for instance, the museum's website and social media were found to be mostly informational and not yet fully utilizing digital resources for outreach. This illustrates how institutional scale determines digital capacity: the British Museum can employ specialist IT staff, mass data storage, and sophisticated search algorithms to host millions of records, whereas small museums focus on minimal online presence and might utilize aggregators or shared platforms to facilitate increased reach.

In addition to sheer numbers, the British Museum's scale advantage also brings obligations. With millions of records, the Museum must ensure high data quality and address issues like duplicate entries or incomplete metadata that might be less prevalent in smaller collections. It also means usage of the Museum's data is global and high-volume – scholars, educators, and the public make heavy demands on the APIs and websites. The Museum invests accordingly in infrastructure robustness (servers, cloud delivery, etc.) to handle this traffic, something a small museum would not typically need to do. Conversely, the impact of digital initiatives at the British Museum can be proportionally large: a new feature on Collection Online could potentially benefit millions of users and set standards that smaller institutions might follow. In summary,

the British Museum's data scale – millions of objects – sets it apart from most peers and aggregators, framing both the ambition and the complexity of its digital initiatives. By contrasting it with a national aggregator like SearchCulture.gr and a local museum like CVAR, we see how scale shapes the possibilities and challenges of museum digital infrastructure in practice.

Chapter 4 Data Management and Strategic Use

This chapter examines how the British Museum moves from securely storing its vast digital resources to using them meaningfully for visitors, researchers, and global communities. Section 4.1 explores long-term data storage and preservation strategies, including the use of semantic standards, linked data, and emerging platforms relevant to large-scale cultural institutions. Section 4.2 discusses how this data becomes “active” through digital exhibitions, AI-driven personalization, and educational initiatives aligned with the Museum’s teaching mission. Section 4.3 addresses the challenges of privacy, policy, and cultural responsibility, stressing the need for governance frameworks that balance innovation with ethical and legal obligations. Together, these sections show how the British Museum treats digital data not only as a technical asset but also as a cultural trust.

4.1 Data Storage, Access, and Preservation Strategies

The British Museum manages one of the largest cultural datasets in the world, covering digitized collections (images, 3D scans, archival documents), conservation records, metadata, research outputs, and large quantities of visitor interaction data. To manage this enormous information ecosystem, the Museum has adopted long-term preservation strategies that ensure data remains both accessible and meaningful despite changing technologies.

A central feature of this strategy is the adoption of semantic standards such as CIDOC-CRM, which allow the Museum to represent collection data in a structured, interoperable way. By using ontologies and controlled vocabularies, the Museum ensures that its records are preserved with their contextual meaning intact and are compatible with international data ecosystems. This approach reflects broader trends, such as Nikolaidou’s semantic documentation framework using CIDOC-CRM and the Getty Art & Architecture Thesaurus, which “promotes a shared understanding and ensures the long-term preservation” of documentation [Nik24]. The MBD2024 synthesis similarly stresses that interoperable infrastructures – shared ontologies, knowledge graphs, linked data – are essential for sustainable curation of cultural data.

In line with this, the British Museum has been expanding its use of Linked Open Data (LOD) to connect its collections with external research resources. Knowledge graphs enable cross-collection analysis, semantic enrichment, and more sophisticated retrieval

of cultural information. This semantic interoperability not only enables present-day access but also future-proofs the Museum's digital assets, ensuring that data remains readable, transferable, and interpretable across generations.

Beyond the semantic layer, the Museum combines on-site repositories with scalable cloud infrastructure, following robust backup and redundancy procedures. Consistent with digital preservation frameworks such as OAIS, the Museum stores digital files in durable, non-proprietary formats and enriches them with detailed metadata. Standards such as Dublin Core, METS, Europeana EDM, and persistent identifiers (DOIs, ARKs) are increasingly integrated to maintain traceability, authenticity, and long-term accessibility.

Scalability is another important consideration. As the British Museum continues to produce "big data"—including high-resolution imaging, 3D modelling, sensor and environmental monitoring data, and user-generated material—its systems must adapt accordingly. Many institutions now collaborate with national aggregators such as SearchCulture.gr; although the British Museum is not part of that platform, the example illustrates how centralized semantic infrastructures can unify heterogeneous datasets at national level. The Museum's ongoing digital partnership projects similarly aim to consolidate dispersed data sources into unified knowledge environments.

Another strategic dimension involves defining what data should be openly accessible. The British Museum embraces an "open access" philosophy for much of its collection metadata and images, contributing to international platforms such as Europeana and Wikimedia Commons. As Jill Cousins highlights, open-data initiatives require a shared vision, collaboration, and adherence to common standards and interoperability principles. At the same time, the Museum manages internal digital records—such as sensitive provenance files—and restricts access where needed for legal, ethical, or conservation reasons. A tiered access model ensures that public engagement and openness coexist with responsible stewardship.

Innovative digital platforms also play a role in the Museum's preservation strategy. Although ARTES is not used by the British Museum, it offers a relevant example of ontology-driven documentation and digital twin technologies that could inform future heritage-sector practices. ARTES creates "distinctive digital twins of artworks and guarantees data protection," supporting provenance tracking, valuation, and conservation documentation [KABH24]. Koutoupas et al. highlight how such platforms demonstrate "unmatched digital documentation and conservation capabilities,"

underscoring how advanced technological solutions can enhance long-term preservation of cultural heritage.

Ensuring discoverability is another core concern. The British Museum seeks not only to preserve data but also to make it visible and intelligible to online audiences. Techniques such as Schema.org markup and JSON-LD annotations, demonstrated by Angelaki et al. [APK24], significantly improve how cultural collections appear in search engines. By enriching the Museum's website with machine-readable semantic descriptions, digital collections become easier for global users to find. The MBD2024 preface further noted work on digital accessibility for big cultural datasets [PD24], highlighting the importance of structured metadata and accessible web interfaces in democratizing cultural knowledge.

Above all, digital preservation at the British Museum is an ongoing, active process. As Chagas et al. argue, digital curation – which includes periodic data migration, checksum audits, and continuous documentation – is essential for sustaining long-term access. The Museum's digital strategy therefore emphasizes not just technology but also organizational readiness: trained staff, sustained funding, policy frameworks, and collaboration with international networks. As Papaioannou and Damigos [PD24] remind us, museums face “significant challenges related to interoperability, data quality, privacy... and ethical stewardship” in the era of big data. Implementing standards, maintaining scalable infrastructure, and formalizing preservation workflows are all essential to meeting these challenges.

In summary, the British Museum adopts an integrated, multi-layered approach to data preservation – using semantic ontologies, linked data, durable formats, standardized metadata, and scalable technologies. Emerging models like digital-twin platforms complement these efforts, while open-access initiatives and web-friendly metadata ensure global reach. These strategies reflect the Museum's commitment to sustaining digital heritage in a trustworthy, accessible and future-oriented form.

4.2 Using Data for Engagement, Personalization, and Learning

The British Museum strategically uses its extensive digital data ecosystem to enhance visitor engagement, enable personalized experiences, and expand its educational mission beyond the physical building. As one of the world's most visited cultural institutions, the Museum has increasingly turned to digital tools, immersive technologies, and data-

driven systems to present its collections in new, interactive, and accessible ways. The following subsections explore how digital exhibits, AI-driven personalization, and educational technologies activate the Museum's data for public benefit, while remaining grounded in scholarly, ethical, and interpretive values.

4.2.1 Digital Exhibits / Interactive Experiences

The British Museum has increasingly adopted visualization technologies – VR, AR, and mixed-reality tools – to bring cultural heritage to life in ways that go beyond static gallery displays. As highlighted at MBD2024, immersive technologies (XR) are becoming essential “instruments for museum display” and “tools for audience engagement, enabling new modes of interpretation, connection, and cultural literacy.” The British Museum's digital strategy reflects this shift, using immersion to help visitors contextualize its vast global collections.

- **Immersive reconstructions and digital environments**

Like the VR reconstruction of the Neolithic settlement of Khirokitia described by Artopoulos et al. [ALD24], the British Museum has developed virtual tours and immersive historical scenes linked to its permanent collections. For example, VR experiences tied to the Museum's ancient civilizations collections – such as virtual explorations of Egyptian tombs or Mesopotamian temples – allow visitors to situate artifacts within their original architectural and cultural environments. These immersive tools mirror the academic-pedagogical dual purpose demonstrated in the Khirokitia project: enabling specialists to test interpretations, and offering visitors a form of contextualized learning that would otherwise be inaccessible.

Similarly, the Museum's AR-enabled mobile applications allow visitors to visualize how objects originally appeared – such as the restored colors of statues, missing architectural fragments, or digitally reconstructed frescoes. These applications reflect the broader research showing AR as an effective way to animate static displays and “bring artifacts out of the display case.”

Enhancing understanding through spatial and multisensory learning, Studies have shown that VR can improve cognitive processing of complex historical information, a finding the Museum has incorporated into its design of digital interpretation tools. By engaging multiple senses – sight, sound, and sometimes haptic feedback – immersive displays align with different learning styles and strengthen knowledge retention.

The Museum's adoption of VR/AR in temporary exhibitions follows the growing international trend noted in the literature, where "large museums with international presence have embraced the use of immersive technologies...in temporary exhibitions." These installations often combine physical and digital storytelling, an approach that resonates with younger audiences and supports accessibility for users with mobility or visual needs.

- **Interactive multi-user experiences**

The British Museum has also explored interactive projections, gesture-based displays, and multi-user environments. By enabling collaborative exploration—similar to the "CAVE-like" interactions described in immersive archaeology projects—these installations position visitors as active participants. Such engagement deepens emotional and intellectual connections between the public and the Museum's global collections.

- **Digital exhibitions as educational and accessibility tools**

Immersive exhibits do not replace physical objects; rather, they enrich their interpretation. VR may offer access to gallery spaces for remote or mobility-restricted audiences, while AR overlay tools act as interpretive layers that bridge the gap between object and context. These uses align with the Museum's mission to broaden access and democratize cultural knowledge.

In conclusion, the British Museum uses immersive data-driven technologies to create meaningful digital exhibits, paralleling examples like Khirokitia [ALD24], and demonstrating how cultural institutions can activate their data to offer transformative visitor experiences.

4.2.2 Visitor Personalization / AI

The British Museum is exploring personalized and adaptive digital systems that tailor the museum experience to individual visitors. While still emerging sector-wide, AI-driven personalization aligns with broader trends identified in MBD2024 and related research.

- **AI chatbots and conversational interpretation**

Building on pilots in other museums, the British Museum is assessing the potential of generative AI chatbots trained on its own digitized collections. Such systems could allow

visitors to ask questions about objects and receive dynamically generated explanations in multiple languages – an approach that mirrors the possibilities described in the MBD2024 reflections on multimodal and generative AI.

These conversational agents combine large language models with the Museum’s metadata and knowledge graphs, enabling context-sensitive narrative responses that adapt to each visitor’s interests and level of expertise.

- **Personalized tours and recommendation systems**

Although the British Museum has not implemented the INDIANA platform directly, the concept parallels the Museum’s ongoing work on dynamic recommendation engines. INDIANA’s use of real-time, context-aware suggestions based on sensor and preference data [MFD+24] illustrates how museums could adapt similar systems to create individualized routes through large collections.

In a British Museum context, such systems could recommend object highlights based on previously viewed items, adjust narrative depth depending on visitor preferences, provide breaks or accessibility reminders based on mobility data or suggest temporary exhibitions aligned with user interests. Real-time adaptation models like those presented by Deliyannis et al. [PD24] are particularly relevant to the Museum’s ambition to develop audio guides or mobile interfaces that adjust narration length, complexity, or topics on the fly.

- **Adaptive exhibits and smart galleries**

The British Museum has also experimented with interactive displays that adjust content according to visitor engagement patterns. Drawing on techniques described in AI-driven behavioral adaptation systems – such as facial-expression recognition or interaction-based difficulty scaling – future galleries may respond in real time to visitor interest or confusion, tailoring content to different learning speeds and styles.

Like all UK institutions, the British Museum must comply with GDPR. Therefore, any personalization strategy must incorporate privacy-by-design principles, pseudonymized or on-device data processing, transparency about data use and user consent.

These constraints echo concerns raised in MBD2024 regarding bias, fairness, autonomy, and responsible AI deployment. The Museum has already begun drafting internal

guidelines for ethical AI use that parallel the “ethical reasoning frameworks” discussed in the conference proceedings.

While AI can enrich the visitor journey, the Museum must avoid over-optimization based on popularity or algorithmic preference. As the literature warns, personalization may inadvertently create “filter bubbles.” For this reason, British Museum personalization tools must support exploration and maintain the curatorial narrative rather than reduce the visit to a data-driven sequence.

4.2.3 Educational Applications / Digital Tools

The British Museum’s educational mission is global, and digital data has become one of its most powerful assets for teaching and outreach. The following examples from the literature illustrate strategies that align closely with the Museum’s own digital learning initiatives.

Projects such as the Goiás Memory Center’s digital curation model [CFG24] highlight the need to transform raw digitized material into structured, pedagogically meaningful resources. The British Museum follows similar approaches through its Schools Programme, online teacher resources, and thematic digital discovery tools. Like the Goiás initiative, the Museum recognizes that teachers need curated sets of objects – organized by theme, period, or curriculum requirements – rather than large, unfiltered datasets.

The Folk Tale Museum scenario [Pap24] demonstrates how big data can uncover patterns in intangible heritage. While the British Museum primarily manages material culture, its vast textual archives and research datasets allow it to support similar data-driven storytelling. For example, digital humanities projects connected to the Museum use aggregated texts, archival materials, and object metadata to explore cross-cultural themes, migration histories, and diasporic narratives.

Such computational approaches to heritage echo the scenario’s emphasis on “Big Data in exhibition design, artistic enhancement and management,” and the British Museum increasingly partners with universities on related digital storytelling projects.

Chalkia, Douka & Sfyridou’s work on digital literary production illustrates how large cultural datasets foster nonlinear reading and narrative creation. The British Museum applies similar ideas through its student programs that use digitized archives, 3D models, and image datasets for creative writing, historical fiction workshops, and digital art

projects. These activities enhance digital literacy and interpretive thinking – mirroring Chalkia et al.’s conclusions about the transformative role of big data in learning.

The CVAR case study [CVAR24] shows how even modest institutions can extend learning through digital channels. For the British Museum, this example reinforces the importance of maintaining a strategic digital presence that is not merely informative but pedagogical. The Museum’s social media storytelling, micro-learning posts, and virtual galleries similarly aim to widen participation and reach learners worldwide.

MBD2024 highlighted several GLAM innovation models (e.g., Dobрева & Papaioannou, Aggeletaki & Mavroudi), emphasizing collaboration, sprint-books, hackathons, and co-creation. The British Museum actively engages in similar initiatives with universities, schools, and cultural organizations, using open data to stimulate creative educational projects, student hackathons, and interdisciplinary digital scholarship.

The Museum’s online courses, object-centered MOOCs, virtual tours, and interactive learning platforms demonstrate how digital data supports global accessibility, multilingual content, and inclusive design – all essential to the Museum’s mandate to broaden cultural literacy.

4.3 Challenges: Privacy, Policy, and Cultural Responsibility

As the British Museum expands its use of digital data, immersive technologies, and AI-driven systems, it confronts a unique combination of opportunities and complex responsibilities. More than many museums, the British Museum operates at the intersection of global cultural heritage, public accountability, and international political scrutiny. This makes the ethical, legal, and cultural implications of data use especially significant. The Museum must navigate challenges related to GDPR and privacy, responsible AI deployment, contested cultural data, international heritage policy, and the stewardship of sensitive digital material.

The following section reframes the challenges discussed in your original text – privacy, policy, and cultural responsibility – through the lens of the British Museum’s realities.

4.3.1 Privacy, GDPR Compliance, and Ethical Data Collection

Like all major UK cultural institutions, the British Museum must comply with the EU/UK GDPR framework, which governs personal data collection, storage, and use. While

museums were once anonymous spaces, the British Museum now manages substantial digital information:

- online ticketing and membership records,
- email subscriptions,
- visitor analytics from the Wi-Fi network,
- interactions with mobile apps and digital guides,
- and data generated from interactive or sensor-based galleries.

As Pediaditaki notes [Ped24], GDPR introduces strict requirements such as **informed consent**, **data minimization**, **storage limitation**, and the **right to be forgotten**, which directly influence how museums design digital visitor experiences.

For the British Museum, this means that every personalization feature, location-based service, or AI-enabled guide must first pass a rigorous series of questions:

- Do we genuinely need this visitor data?
- Can we achieve this goal anonymously?
- How will the Museum communicate consent transparently?
- Can the system function with privacy-by-design principles?

These issues are heightened at a museum of the British Museum's scale, where millions of visitors per year generate vast potential datasets. Staff must be trained in GDPR compliance, and the Museum maintains a dedicated data protection function, mirroring what Pediaditaki describes as the sector's growing need for **internal expertise in data security and ethical data management**.

While CCTV is indispensable for collection security, using video or sensor data for behavioral analytics raises ethical concerns. Museum Big Data reflections in MBD2024 emphasize "personal data, privacy, data protection regulations (such as GDPR), transparency, and accountability" as central issues.

For the British Museum, deploying technologies like Bluetooth beacons, heat-mapping, or AI-based emotion analysis is especially sensitive because the Museum serves a global and diverse public. Visitors may come from regions with different privacy expectations.

The Museum positions itself as a trusted civic institution.

Therefore, the Museum tends to favor **aggregate, anonymized analytics**—e.g., measuring overall crowd flow rather than tracking identifiable individuals. Technologies such as facial recognition are avoided for visitor studies due to compliance risks and ethical limitations.

4.3.2 Responsible and Culturally Sensitive Use of AI

Introducing AI tools—whether for visitor services, interpretation, or collection management—brings ethical considerations beyond privacy. AI can unintentionally reinforce cultural biases, misinterpret heritage objects, or blur the distinction between scholarly fact and algorithmic conjecture.

AI accuracy and interpretive authority

The British Museum holds one of the world’s most symbolically and politically charged collections. If an AI model mislabels artifacts or offers historically inaccurate information, it risks reinforcing stereotypes or spreading misinformation about world cultures.

This reflects concerns from MBD2024, where scholars caution that AI may jeopardize “human interpretive nuance and authority.” For the British Museum, this is especially critical because its exhibitions often relate to colonization, contested heritage, and global narratives. Any generative AI used in galleries or digital tools must therefore be trained using vetted museum metadata, undergo human review by curators, clearly disclose when content is AI-generated and avoid automated statements that could appear authoritative without scholarly validation. For instance AI trained on generic datasets may interpret African, Indigenous, Middle Eastern, or Asian artifacts through Western-coded assumptions. The British Museum must actively prevent such distortions, a challenge aligned with the MBD2024 emphasis on “inclusive and culturally sensitive approaches to museum innovation” [PD24].

To mitigate this, the Museum incorporates culturally aware metadata created with input from researchers and, in some cases, source communities manual oversight for automated descriptions and restrictions on generative AI use in sensitive or contested areas.

4.3.3. Cultural Stewardship and Sensitive Digital Heritage

The British Museum faces stronger scrutiny than most institutions because many of its objects originate from former colonies, Indigenous communities, and regions affected by conflict. Digitizing such collections and distributing data globally introduces unique ethical responsibilities. Source communities may have cultural protocols around: images of ancestors, sacred objects, funerary items or materials associated with colonial extraction.

While the Museum maintains an open-access digital policy for much of its collection, it must ensure that openness does not violate cultural norms. This aligns with the text's discussion of culturally sensitive data governance. In practice, the British Museum sometimes restricts access to certain images, adds contextual warnings, consults with descendant or source communities or provides mediated access rather than unrestricted downloads. These decisions reflect a shift toward shared authority over cultural data.

4.3.4 International Policy, Repatriation Debates, and Digital Data

British Museum data governance is deeply shaped by international policy frameworks. As Chatzidi's invited paper at MBD2024 demonstrates, "the role of museums (international vs. national) on the increasing scale of the trade of cultural objects" creates tensions among different legal and ethical regimes. For the British Museum, this has concrete implications. Some countries restrict the digitization of archaeological material. Publishing 3D models may inadvertently facilitate illicit reproduction. Certain artifacts are subjects of ongoing repatriation negotiations (e.g., Parthenon Sculptures, Benin Bronzes), making their digital surrogates politically sensitive. UNESCO conventions and national heritage laws may limit what digital data can circulate freely online. Thus, even open-access initiatives must be weighed against international cultural property laws and diplomatic considerations.

While the British Museum is committed to open cultural data, it must respect copyright in modern works, image rights, donor agreements and restrictions on sensitive objects. This echoes the broader policy tension described in your original text: balancing openness with legal and ethical constraints.

4.3.5. Cybersecurity and Protection of Digital Heritage

Although not extensively discussed in your prompt, cyber-risk is a real issue for the British Museum. The Museum holds personal data from millions of visitors and donors. It also has high-resolution images of priceless artifacts, internal research records, and sensitive provenance files. A breach would compromise public trust and potentially expose culturally sensitive information. Therefore, the Museum invests in strong cybersecurity, backups, and disaster-recovery systems—an integral part of responsible digital stewardship.

4.3.6 Cross-Institutional Collaboration and Policy Alignment

As global museums increasingly collaborate on joint digital platforms, the British Museum must navigate differences in privacy legislation, open-data policies, cultural property law and community expectations. This echoes your discussion of MOUs and transnational GLAM frameworks. European and international GLAM labs are working toward common models—efforts in which the British Museum participates. Aligning digital protocols across borders ensures that shared data respects all partners' legal and ethical requirements.

For the British Museum, the growth of museum big data brings exciting possibilities but also heightened responsibilities. GDPR compliance, privacy-by-design, and cybersecurity form the legal foundation. Ethical AI, cultural sensitivity, and responsible interpretation ensure the Museum remains a trusted authority. International policy complexities—highlighted by Chatzidi's analysis—require careful navigation of cultural property norms, digital repatriation issues, and divergent national laws. Ultimately, innovation and ethical stewardship must progress together. By embedding cultural responsibility, transparency, and consultation into every digital initiative, the British Museum can continue to lead globally while honoring its obligations to communities, visitors, scholars, and the cultural heritage it preserves.

Chapter 5 – Conclusion: Results and Future Directions

This chapter distills what the study has uncovered and where the field should head next. Section 5.1 synthesizes the main results: the scale and diversity of museum data, the role of interoperable platforms and enriched metadata, and the emerging potential for ethical, data-driven personalization—illustrated through the British Museum case. Section 5.2 turns these findings into a research agenda, highlighting priorities in AI and machine learning, immersive interpretation, stronger knowledge graphs and standards, and inclusive governance. Section 5.3 closes with practical takeaways for institutions of all sizes, arguing that robust data strategy is now core to collection care, public access, and learning. Together, these sections link evidence to action, setting out a roadmap for museums to steward data as a cultural asset and a public good.

5.1 Results and Contributions

This research polled how museums manage and utilize their data, utilizing the British Museum as the key case study in conjunction with standard industry approaches. The findings stipulate the range and extent of museum data, systems and platforms utilized to collect and store such data, and what potential exists for tailored experiences if data are utilized to their full extent. Key findings and contributions are:

Data Diversity and Volume: Museums now manage enormous, diverse datasets in their collections and activities. For example, the Greek national aggregator SearchCulture.gr consolidates over 860,000 records from 94 providers (museums, archives, galleries, etc.) on archaeology, history, fine arts, and intangible heritage. This indicates both the sheer volume of cultural heritage data as well as its diversity in content types and subject domains. It is “no surprise that museums have tremendous amounts of data,” and recent strides have focused on structuring collection information and making it accessible to the public. The British Museum’s own collection (around 8 million objects) exemplifies this scale – a volume of data that poses challenges for documentation but also offers unprecedented opportunities for analysis and engagement.

Cross-Platform Data Management: Museum data is dispersed across different platforms and systems. Across museums, one finds collection databases and digital catalogues, whereas in external channels, one has data sharing via websites, open data portals, and

aggregators. Rendering these platforms interoperable has been a major issue. Initiatives like SearchCulture.gr enforce common standards (e.g. Europeana Data Model and linked open vocabularies) to normalize metadata of heterogeneous collections. Similarly, many museums contribute to international platforms (such as Europeana or the upcoming Museum Data Service) to **centralize access** to cultural data. The British Museum has been a leader in cross-platform engagement – partnering with tech companies and adopting open data practices – thereby demonstrating how a large institution can act as a data hub. The use of standard formats and APIs means that museum data can flow between institutional databases, national aggregators, and global knowledge graphs, mitigating silos and enabling broader reuse of cultural information.

Metadata Enrichment and Findability: A key contribution of this work is underscoring the importance of **high-quality metadata** for museum collections. Rich, standardized metadata enhances findability and usability of cultural content. For instance, semantic enrichment (as implemented by SearchCulture.gr) adds consistent tags for object types, subjects, persons, places, and periods, answering the fundamental “who/what/when/where” questions for each item. The result is more precise search results and new browsing functionalities (like map-based discovery and thematic collections) built on top of big data. This research highlights that investing in metadata curation – through controlled vocabularies, ontologies (e.g. CIDOC-CRM), and linked data – directly contributes to better data integration and user discovery.

Data quality (accuracy, consistency, up-to-date records) is as crucial as quantity: without clean and connected metadata, the sheer volume of museum data cannot be effectively exploited.

Personalization Potential: Perhaps the most forward-looking finding is the significant potential for **personalized experiences** in museums, powered by big data. As collection data becomes more structured and interoperable, museums can deploy recommendation systems and tailored content delivery. For example, embedding rich structured data (using standards like Schema.org/JSON-LD) into museum websites enables sophisticated search and filtering, and even personalized content recommendations based on a visitor’s interests or browsing history. The value of such **personalized engagement** is evident – users can receive suggestions for exhibits or artifacts aligned with their preferences, creating a more interactive and relevant museum experience. The groundwork for personalization is visible in current practices: some museums analyze visitor data (e.g. the paths people take through exhibits) to inform exhibit design, and

others use online interaction data to recommend related objects or stories. Our research contributes by mapping how personal data, if managed ethically, could be integrated with collection data to enhance learning and enjoyment.

These results collectively show that proper management of museum data – from comprehensive digitization and metadata standardization to cross-platform sharing – unlocks new value for both museums and the public. The British Museum’s case exemplifies these trends. It has leveraged big data analytics to holistically understand visitor behaviors in its galleries, partnering with Microsoft to map how visitors move through 800,000 square feet of gallery space and identifying which exhibits draw attention or are bypassed. Insights from such analyses have helped improve gallery layouts and interpretation strategies. Moreover, in response to recent challenges, the British Museum committed to fully document and digitize its entire collection – a massive undertaking that underscores the importance of data transparency and accessibility. By setting out to digitize and publish an eight-million-item catalogue, the British Museum is “*setting a leading example*” for comprehensive data stewardship, one that other institutions are now encouraged to follow [Tap24]. In sum, this research’s contributions lie in articulating how data types, volume, platforms, and personalization intersect in the museum context, and in highlighting best practices (exemplified by initiatives at the British Museum and elsewhere) that can guide cultural institutions worldwide.

5.2 Future Research Directions

Based on the research above, several prospects emerge for future research and development in museum data management. The museum industry is on the threshold of an era of revolutionary transformation by technology and shifting public ambitions. Top priority areas are:

Artificial Intelligence and Machine Learning: AI being integrated into museum data strategies is an exciting new front. Sophisticated AI techniques – e.g. multimodal large language models and vision-language models – can be combined with structured cultural datasets (e.g. museum knowledge graphs) to unlock new potential. For instance, machine learning may be applied to automatically classify objects, detect patterns in visitor movement, or even generate narrative object descriptions. Museums are also looking to AI for two general purposes: it can help them learn more about their audience and offer ways to develop programs better suited to what visitors are searching. Future research must explore AI-driven personalization (e.g., virtual guides or chatbots that tailor tours

to visitor interests) and predictive analytics for collections (e.g., predicting trends in exhibitions or predicting vulnerable artifacts with sensor data). However, it is important to have clear methodological frameworks and maintain curatorship control – AI should augment, and not replace, curatorial and pedagogic knowledge. The development of explainable and responsible AI in museums will be an integral area of research, making algorithmic recommendations clear and in line with museum ethics.

Immersive and Interactive Experiences: Immersive technologies like Virtual Reality (VR), Augmented Reality (AR), and interactive media are revolutionizing the way that museums interact their audiences with museum content. During the last few years, large museums in general have embraced immersive tech in their operations, particularly in temporary or special exhibitions. These technologies offer human-centered, real-world experiences of cultural information that allow visitors to walk virtually into historical environments or view artifacts in 3D detail. Future research will likely focus on expanding the affordability and availability of immersion for museums of all sizes. VR also has potential in scholarship and conservation, for example, archaeologists using VR reconstructions to try out hypotheses about site layouts. With cheaper hardware, museums can experiment with in situ VR installations, AR apps on smartphones for independent visits, and even fully virtual museums on the internet. Research should also examine how these experiences impact learning results and how to implement immersion effectively without overpowering the actual objects. The synergy of big data and immersion (e.g. using data to dynamically adjust an AR tour to a visitor’s behavior) could be a rich area for innovation.

Enhanced Metadata and Knowledge Networks: The future of museum informatics lies in ever better metadata and connected data systems. Metadata automation will become important – machine learning can assist in tagging images, identifying depicted subjects, or translating catalog records, thereby reducing the manual metadata backlog. Co-designed tools (user-friendly curation interfaces, validation dashboards, etc.) are needed to help museum staff work alongside these automated systems. Moreover, expanding the use of ontologies and knowledge graphs will enable museums to link their data with external datasets (biographical data, geographic data, historical events), providing richer context for collections. Research should continue on establishing cross-institution collaborative infrastructures for data exchange. Platforms like ARTES (for art collections) and national aggregators like SearchCulture.gr demonstrate the benefit of sharing data and technical solutions across museums. By pooling resources and adopting common standards, even smaller institutions can partake in a “big data” ecosystem rather than

being left behind. Future investigations might also look at real-time data integration – for example, connecting live museum attendance data with collection information to dynamically recommend less-crowded galleries or related exhibits, enhancing visitor flow and experience. In short, better metadata, interoperability, and data-sharing protocols will be foundational to all other digital innovations in museums.

Inclusive and Ethical Data Practices: As museums become more data-driven, there is a pressing need to address the ethical and inclusive dimensions of this shift. Future research should examine frameworks for data governance that uphold privacy, consent, and cultural sensitivity. Museums must develop policies for responsibly collecting and using visitor data (e.g. respecting anonymity in analytics, obtaining consent for location tracking in galleries) and for handling culturally sensitive collection data (such as sacred objects or human remains) in digital platforms. Ethical AI is another critical area – ensuring algorithms do not introduce bias or erode diversity in the content shown to visitors. There is also a call for improving data literacy among museum professionals: investing in training so staff can critically interpret data and AI outputs. Inclusivity in museum data means not only reaching diverse audiences but also involving community voices in how data is created and presented. For example, engaging source communities in enhancing metadata for indigenous collections can ensure respectful and accurate representation. Furthermore, bridging the digital divide is essential: future initiatives should make advanced museum technologies accessible to differently-abled visitors and to institutions with fewer resources. The democratization of museum big data – where tools and benefits are shared broadly – is a key direction. Strengthening collaborative networks and open-source projects can support smaller museums, ensuring the digital transformation does not only favor well-funded institutions. Ultimately, research must continue to foreground ethical, inclusive, and culturally sensitive approaches so that museum innovation benefits all stakeholders and honors the public trust vested in cultural institutions.

5.3 Concluding Remarks

The evolving practices highlighted in this thesis all point to one central insight: proper museum data management is now fundamental to museum success and community value. Museums like the British Museum serve as examples in this space. The British Museum's high-profile digitization cataloguing project – undertaking to digitize and publish its entire collection of some 8 million objects – demonstrates leadership in openness, access, and accountability. Not only does this work safeguard the collection

with better documentation but also greatly facilitate public accessibility, allowing anyone anywhere in the world to search and research the museum's holdings. By so doing, the British Museum is not merely preserving the past but empowering future scholarship and engagement.

Good data management, of course, is as much about high-tech and big data programs as it is about alignment with core museum values. At their essence, museums are about acquiring, conserving, interpreting, and sharing human heritage. Durable data practices enrich each of these needs: from online preservation of artifacts and rigorous provenance documentation, to data-intensive interpretation that has the capability to reveal hidden connections between objects, to sharing knowledge through open data and web platforms. The results of this study reinforce that museums must treat their data as a strategic resource. An investment in strategic digital infrastructure, staff training, and institutional cooperation will be paid back with better collection care and improved visitor experiences.

Overall, as museums go forward in the era of big data, those that are adopting holistic, ethical, and forward-thinking data approaches will be better positioned to thrive. The British Museum's innovations in data-driven practice demonstrate what is possible, but all museums, both small and large, can take action to improve its data work. In doing so, museums worldwide will not only become streamlined and effective within, but also fulfill their public mission in the digital age: to inspire, to educate, and to enhance the relationship of citizens with cultural heritage. Tomorrow's museums will be built on foundations laid today in innovation and data management, to save and to bring to life our global heritage through richer and more personal experiences for generations to come.

References

- [And24] Alexandra-Aikaterini Andrianou. Cultural Landscapes and Big Data in Qualitative Research: The Case Study of the Martyr Village of Kommeno, in Arta, Greece. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [AGP+24] Georgia Angelaki, Harris Georgiadis, Agathi Papanoti, Elena Lagoudi. Leveraging Linked Data to develop rich discovery services for big heterogeneous cultural data; the case of the National Cultural Heritage Aggregator SearchCulture.gr. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [ALD24] Georgios Artopoulos, Nicolas Loucas, Odile Daune-Le Brun. An immersive workflow of co-creating virtual reconstructions of archaeological sites as an act of transdisciplinary research. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [AKB24] Avgoustinos Avgousti, Athanasios Koutoupas, Nikolas Bakirtzis. Applied Computing and Big Data: A Case Study of the Pancyprian Gymnasium Historical Archive. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [APK24] Avgoustinos Avgousti, Georgios Papaioannou, Athanasios Koutoupas. Leveraging Big Data and Computing Technologies for Improved Digital Accessibility in Cultural Heritage. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*

-
- [CAM+21] CAMOC: Book of Proceedings – CAMOC Conference 2021, Barcelona. CAMOC. URL: [CAMOC-CONFERENCE-2021-BARCELONA-BOOK-OF-PROCEEDINGS-website.pdf](https://www.camoc-conference-2021-barcelona-book-of-proceedings-website.pdf), 2023.
- [CFG24] Paula Roberta Chagas, Alison Carlos Filgueiras, Feliz Alberto Ribeiro Gouveia. The Goiás Memory Center and heritage education: Reuse of cultural collections and Museum data in schools in a Brazilian country town (Morrinhos / Goiás). *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [CFB+24] Paula Roberta Chagas, Alison Carlos Filgueiras, Juliana Vasconcelos Braga, Feliz Ribeiro Gouveia. Local history education and Museum data: constructing a model for heritage digital curation. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [CDS24] Natassa Chalkia, Areti Dimitra Douka, Eleni Sfyridou. Digital Literary Production: Transformation of Reading, Writing, and Interpretive Skills. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [Cha24] Sofia Chatzidi. The role of the museums (international versus the national) on the increasing scale of the trade of cultural objects. Different agendas and frames of confronting existing policies and data on this trade. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [CC15] Creative Commons: Digital community, heritage and open access. Medium (2015). URL: <https://medium.com/creative-commons-we-like-to-share/digital-community-heritage-and-open-access-7e3477ef2131>
-

- [DMP+24] Ioannis Deliyannis, Ioanna Afroditi Mazi, Sofia Maria Poulimenou, Despina Elisabeth Filippidou, Eleni Christodouloupoulou, Anastasios Manos, Minas Pergantis. AI-Driven Personalization in Tourism: Enhancing Visitor Experiences Through Real-Time Data. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [PD24] Papaioannou Georgios, Matthew Damigos. Museum Big Data: Emerging Practices, Challenges, and Opportunities in Museum Contexts – A Preface and a Synthesis of the MBD2024 Conference. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*
- [Gia11] Elisa Giaccardi. Heritage and social media: Understanding heritage in a participatory culture, 2013. URL: [\(PDF\) Heritage and Social Media: Understanding heritage in a participatory culture.](#)
- [Gia12] Elisa Giaccardi. Heritage and Social Media: Understanding heritage in a participatory culture. *OverDrive*, 2012. URL: [eBook - Heritage and Social Media by Elisa Giaccardi · OverDrive: Free ebooks, audiobooks & movies from your library.](#)
- [Kit14] Rob Kitchin. The data revolution: Big data, open data, data infrastructures and their consequences. *SAGE Publications, Thousand Oaks*, 2014.
- [KAB+24] Athanasios Koutoupas, Avgoustinos Avgousti, Nikolas Bakirtzis, Sorin Hermon. ARTES: Innovating Art Management with Big Data and Digital Twin Solutions. *In Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024.*

- [Lin15] Matthew Lincoln. Unsustainable museum data. *Matthew Lincoln Blog*, 2015. URL: [Unsustainable museum data](#)
- [MFD+24] Anastasios Manos, Despina Elisabeth Filipidou, Ioannis Deliyannis, Nikolaos Pavlidis, Vasilis Argyros, Ioanna Mazi. INDIANA: Personalized travel recommendations using wearables and AI, 2024. URL: [Personalized travel recommendations using wearables and AI](#)
- [MKT24] Michalis Mountantonakis, Manos Koumakis, Yannis Tzitzikas. Combining LLMs and Hundreds of Knowledge Graphs for Data Enrichment, Validation and Integration: Case Study in the Cultural Heritage Domain. In *Proceedings of MBD2024: International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024*.
- [Tomsk-n.d.] Музей в контексте культуры участия (Museum in the context of participatory culture). Tomsk Regional Museum, Tomsk (n.d.). URL: <https://tomskmuseum.ru/content/editor/Posetitel/OpiFt%20kolleg/Muzej-v-kontekste-kultury-uchastiya.pdf>
- [Nik24] Andriana Nikolaidou. Access to Artists' Studios: An Ontology-Based Framework for the Documentation and Analysis of Artists' Studios. In *Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024*.
- [Pap24] Melina Papadopoulou. Discovering Fairytale Paths: Exhibition Design, Artistic Enhancement, Management and applied data in the conceptual Folk Tale Museum Scenario in Zagori, Greece. In *Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024*.

- [Par13] Ross Parry. The End of the Beginning: Normativity in the Postdigital Museum. *Museum Worlds*, 2013. URL: [The End of the Beginning: Normativity in the Postdigital Museum | Request PDF](#)
- [Ped24] Antonia Pediaditaki. Personal data, big data, cultural heritage and GDPR: realities and challenges. In *Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024*.
- [RA25] Emily Rees-Koerner, Lydia Ackrell. Making museum research more visible: Open Access in the GLAM sector. *Science Museum Group Journal*, Issue 23, Spring 2025. DOI: [Open Access in the GLAM sector](#).
- [Sfy24] Eleni Sfyridou. Online Marketing Tools and Small Arts Organizations: A Case Study on the Centre of Visual Arts and Research (CVAR) in Cyprus. In *Proceedings of MBD2024: 3rd International Conference on Museum Big Data, Athens, Greece, November 18–19, 2024*.
- [Sim10] Nina Simon. The Participatory Museum. 2010. URL: <https://participatorymuseum.org/read/>
- [Tap24] Oliver Tapper. All museums should follow the British Museum’s lead in digitising. 2024. URL: <https://www.globallegalpost.com/news/all-museums-should-follow-the-british-museums-lead-in-digitising-collections-599237447#:~:text=Although%20prompted%20by%20the%20recent,museums%20to%20do%20the%20same>
- [BM-n.d.] The British Museum: About us. The British Museum. URL: <https://www.britishmuseum.org/about-us>

- [VCSR18] Arnold P. O. S. Vermeeren, Licia Calvi, Amalia Sabiescu, Raffaellat Rocchianesi. Future museum experience design: Crowds, ecosystems and novel technologies. *In: Proceedings of the International Conference on Designing Interactive Systems, 2018*. URL: [\(PDF\) Future Museum Experience Design: Crowds, Ecosystems and Novel Technologies](#).
- [WLW10] Günter Waibel, Ralph LeVan, Bruce Washburn. OCLC Research: Museum Data Exchange: Learning How to Share - Final Report to The Andrew W. Mellon Foundation. *OCLC Research Publications, 2010*. URL: [Museum Data Exchange: Learning How to Share](#), 2010.
- [ZF23] Mariana Ziku, Bettina Fabos. Digital Community Heritage and Open Access. *Creative Commons: We Like to Share (blog)*, 2023. URL: [Digital Community Heritage and Open Access](#).

Appendix A: Comparative Metadata Overview of Cultural Heritage Platforms

This appendix presents a comparative overview of three digital cultural heritage platforms – SearchCulture.gr (Greece), the British Museum’s online collection (UK), and the Centre of Visual Arts and Research (CVAR, Cyprus). These platforms were selected to illustrate different approaches to digital heritage management, metadata practices, and levels of technical infrastructure across institutions of varying scale. The comparison focuses on their metadata strategies, interoperability features, and technological choices, emphasizing how museums and archives support access, discovery, and reuse of cultural data in the digital age.

Institution	Platform Name & URL	Number of Digitized Records	Collection Types	Metadata Standards / Vocabularies	Interoperability Features	Technologies / Platforms Used
National Documentation Centre (EKT), Greece	SearchCulture.gr	~860,000 records from 94 providers	Archaeology, folklore, fine arts, manuscripts, historical photos, intangible heritage	Dublin Core (Qualified), Europeana Data Model (EDM), Getty AAT, DCMI terms	APIs, Open Data License (CC-BY), Linked Open Data, Europeana Aggregator	Custom CMS, semantic enrichment services, SPARQL endpoint
The British Museum, UK	British Museum / Collection Online	~4.5 million objects accessible online (from ~8 million total)	Archaeology, ethnography, coins, prints, manuscripts, art, decorative arts	Custom metadata schema aligned with SPECTRUM, internal ontologies, schema	Open Access API, RDF/JSON-LD, CC0 license for metadata and images, Europeana contributor	Custom-built collection management system, Open Data API
Centre of Visual Arts and Research (CVAR), Cyprus	CVAR Online Collection	~20,000 records (estimated)	Paintings, photos, archival documents, audiovisual material, books	Dublin Core, local schema, basic AAT usage	Minimal: website access and catalog search; limited open data exposure	Basic digital archive (no public API); limited metadata export

Table 2 Comparison of Metadata Practices Across Cultural Heritage Platforms

The following table presents an overview of the standard metadata fields provided by SearchCulture.gr. These fields illustrate how museum data is structured to enable interoperability, findability, and reuse in national and European infrastructures.

Field	Description
Title	The official name or title of the cultural object.
Creator	The individual or group responsible for creating the object (artist, author, etc.).
Date	The time period or specific date associated with the object
Type	The object type (e.g., photograph, sculpture, document, artifact).
Subject	Keywords or themes describing the object's content.
Description	A brief narrative providing context or interpretation.
Spatial Coverage	The geographic area related to the object (place of origin, depiction).
Temporal Coverage	The historical period associated with the object.
Institution / Provider	The museum, archive, or library contributing the digital record.
Rights / License	Usage rights and licensing information (e.g., CC-BY, Public Domain).
Identifier	A unique record identifier in the aggregator's database.
Link to Digital Item	A persistent link to the object's digital representation.

Table 3 Core Metadata Fields Used in SearchCulture.gr

The following table illustrates the core metadata fields used in the British Museum's Online Collection. These fields ensure consistent documentation and facilitate access, research, and interoperability with other cultural heritage systems.

Field	Description
Object Number	A unique identifier assigned to each object in the Museum's collection.
Title / Object Name	The name or title given to the object (may include descriptive labels).
Creator / Maker	The artist, craftsman, or culture responsible for creating the object.
Production Date	The date or time range when the object was made.
Place of Production	Geographic location where the object was created.
Materials / Technique	The physical materials and methods used in the object's creation.
Dimensions	Physical measurements of the object (height, width, depth, weight).
Description / Notes	Narrative description and contextual details.
Collection / Department	The Museum department responsible for the object (e.g., Prints & Drawings, Coins & Medals).
Provenance / Acquisition	Information about how the Museum acquired the object.
Subjects / Keywords	Themes, topics, or classifications associated with the object.
Associated People/Places	Links to historical figures, cultures, or locations related to the object.
Rights / License	Usage rights and copyright information (often CC-BY-NC-SA).
Digital Image	Link to digital representation(s) of the object.

Table 4 Core Metadata Field used in the British Museum Online Collection

Screenshots from the SarchCulture.gr platform (Hellenic Aggregator, operated by the National Documentation Centre, EKT)

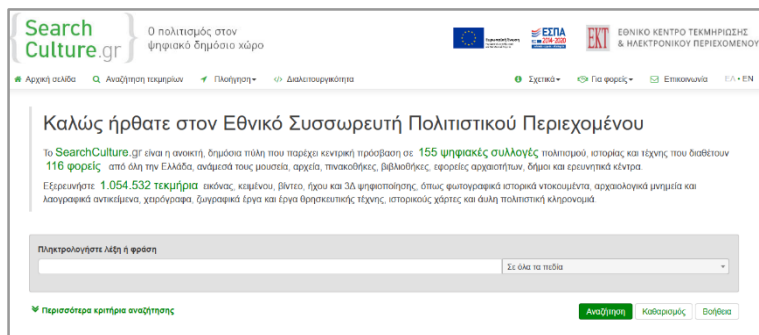


Figure 3 Home page of SearchCulture.gr (National Aggregator of Digital Cultural Content)

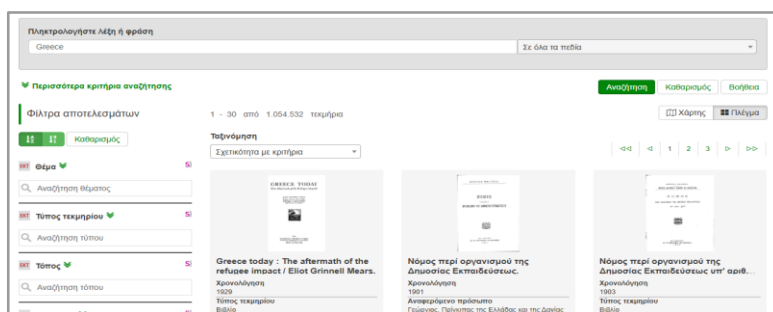


Figure 4 Search interface of SearchCulture.gr



Figure 5 Item detail page for a Greek vase (EKT digital record view)

Screenshots from The British Museum Collections Online portal

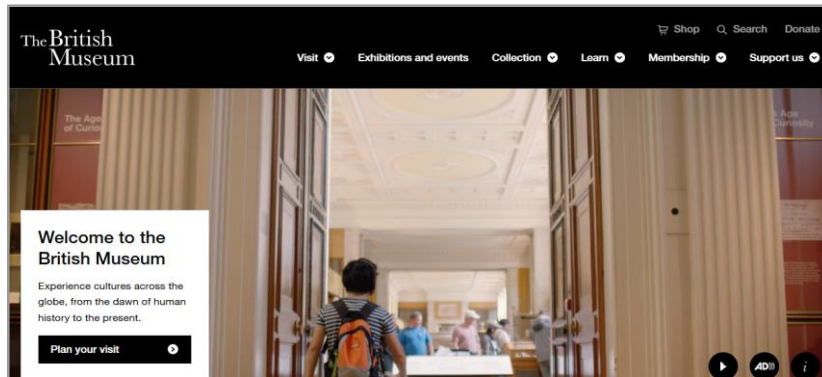


Figure 6 Home page of The British Museum Collections Online portal

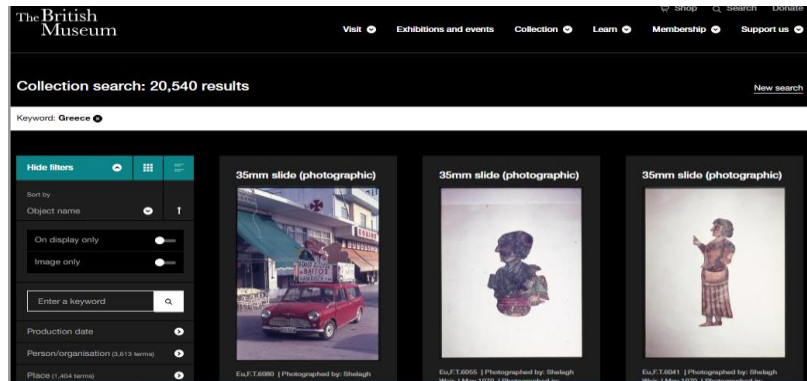


Figure 7 Search Interface of The British Museum/Collections Online

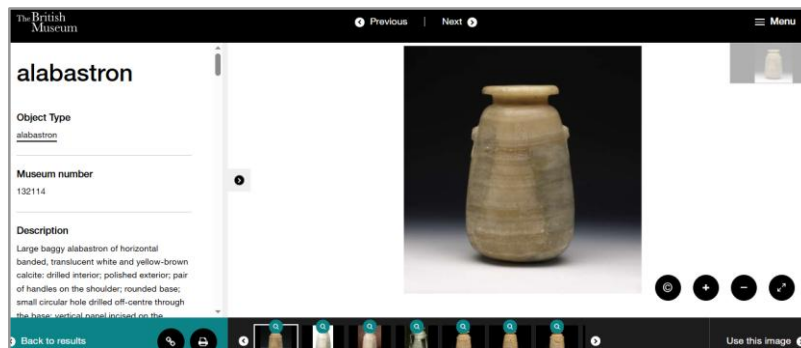


Figure 8 Object record showing metadata fields and image viewer

Appendix B: The British Museum: Digital Infrastructure and Data Framework

The British Museum has developed strong digital systems to manage, preserve, and share its large collections. It uses detailed metadata and applies semantic web tools, open APIs, and knowledge graph methods so its data can connect with other cultural heritage platforms. This appendix gives a short overview of the technical framework behind the Museum's digital work, focusing on how it ensures data compatibility, public access, and modern visualization tools.

Component	Description	Relevance
Database System	Collection database storing ~8 million object records, with ~4 million publicly accessible online.	Core backend for cataloguing, storage, and retrieval.
Metadata Standards	Uses CIDOC-CRM for semantic relationships; Dublin Core and EDM mappings for interoperability.	Ensures data can connect with Europeana and other aggregators.
Linked Open Data (LOD)	Public SPARQL endpoint and RDF dumps available since 2015.	Enables semantic queries, integration with external knowledge graphs.
APIs	RESTful APIs and downloadable datasets for developers and researchers.	Facilitates reuse of BM data in apps, research, and education.
3D & Visualization	Employs photogrammetry and 3D scanning for selected objects; integrated into digital exhibits.	Supports immersive experiences and research reconstructions.
Immersive Tech	Collaboration with Samsung Digital Discovery Centre for VR/AR projects (e.g. Bronze Age roundhouse).	Enhances visitor engagement through virtual environments.
Data Governance	Clear open-access policy for metadata and images (where copyright permits). Tiered access for sensitive data.	Balances transparency with legal/ethical obligations.

Table 5 British Museum – Simplified Digital Infrastructure Diagram

This schematic illustrates the flow of data within the British Museum’s digital ecosystem. The central **Collection Database** holds the museum’s core object records (around 8 million entries). Through **APIs** and a **SPARQL endpoint**, the data becomes accessible to both internal systems and external platforms. These interfaces enable public access via the **Collection Online** portal, collaboration with external partners such as **Google Arts & Culture**, and integration into **VR/AR projects** that reuse collection data for immersive storytelling and education. The diagram highlights how interoperable components – metadata standards, open APIs, and semantic web technologies – allow the Museum to share its collections seamlessly across multiple digital environments.

British Museum - Digital Infrastructure (simple)

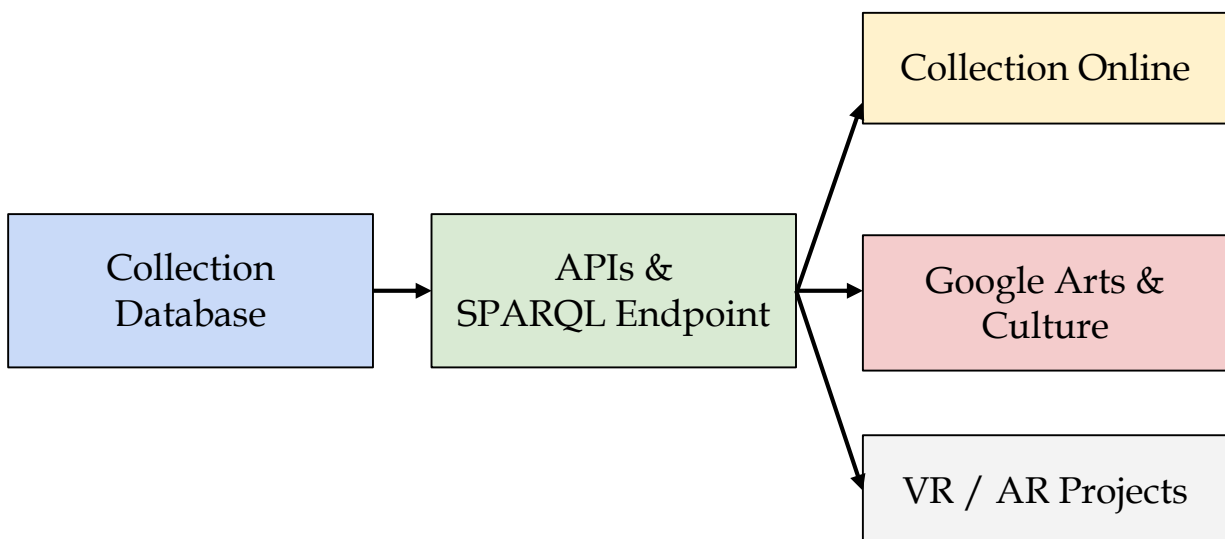


Figure 9 British Museum – Simplified Digital Infrastructure Diagram

This schematic RDF graph represents how the British Museum models relationships between data entities using semantic web principles. The example shows the *Rosetta Stone* (classified as an “Object”) linked to its *Creator* (the *Ptolemaic Dynasty*) through the property **hasCreator**, and to its *Discovery Place* (*Memphis*) through the property **discoveredAt**. Each arrow illustrates a semantic relationship defined in RDF (Resource Description Framework), demonstrating how museum data can be structured for machine-readable connections across knowledge systems.

Minimal RDF graph : British Museum

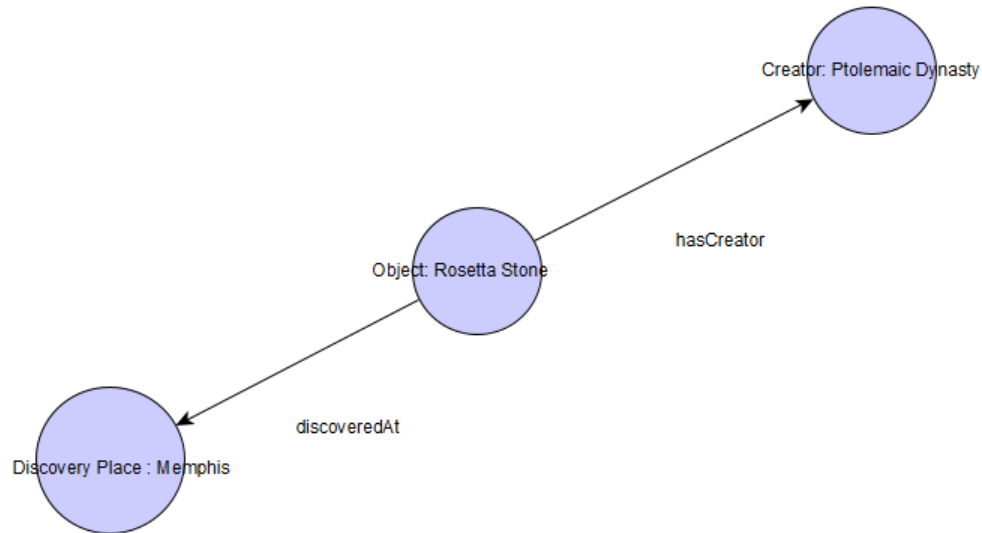


Figure 10 Example RDF triples linking an object to its creator and findspot

This RDF diagram shows how data about the *Rosetta Stone* can be represented in a semantic format using museum ontologies. Each box represents an RDF resource:

- the *Object* (**ex:Object_rosetta**) defined as a **bm:CollectionObject**,
- its *Creator* (**ex:Person_ptolemaic**) defined as a **foaf:Person**, and
- its *Discovery Place* (**ex:Place_memphis**) defined as a **schema:Place**.

The arrows represent RDF predicates that define relationships between these entities:

- **bm:hasCreator** links the object to the *Ptolemaic Dynasty*,
- **bm:foundAt** links it to *Memphis*.

This example illustrates how museum data (objects, people, places) can be expressed as *RDF triples* for interoperability and integration with other cultural heritage datasets.

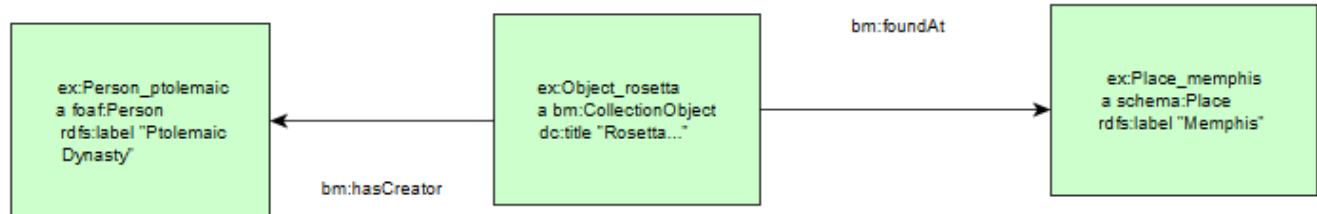


Figure 11 Schematic triple graph: SPARQL query

bm: museum-specific ontology terms (e.g., `bm:CollectionObject`, `bm:hasCreator`, `bm:foundAt`)

dc: Dublin Core (e.g., `dc:title`)

foaf: FOAF (persons/agents)

schema: Schema.org (places, creative works)

rdfs: RDF Schema (labels/types)

Although the British Museum's public SPARQL endpoint is no longer online, the model below reflects common practice in GLAM LOD deployments.

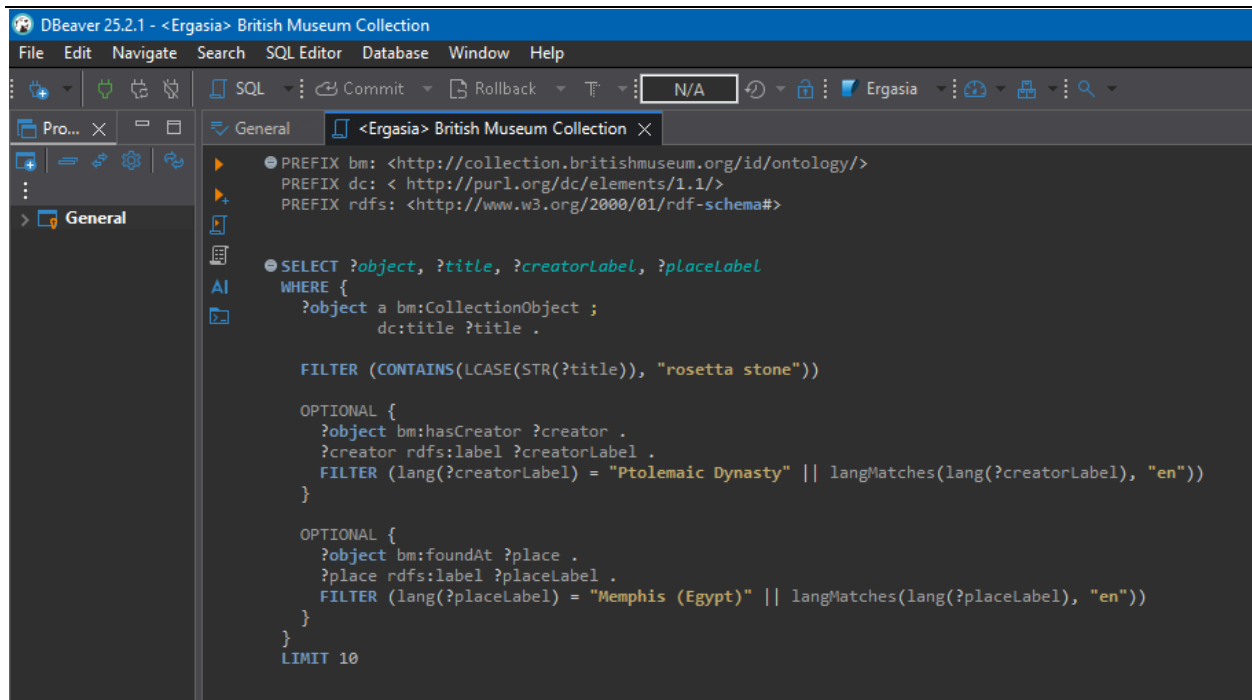


Figure 12 SPARQL query (the equivalent of SQL for RDF/Semantic Web data)

For the British Museum, the former endpoint (which is no longer active) accepted SPARQL queries in an RDF model based on the CIDOC-CRM / Museum Ontology.

This query demonstrates:

- the object "Stela / Inscribed Stone"
- the title "Rosetta Stone",
- the creator "Ptolemaic Dynasty",
- and the findspot "Memphis (Egypt)".

Appendix C: Terminology Index

English Term	Ελληνικός Όρος (Greek Term)
AI (Artificial Intelligence)	Τεχνητή Νοημοσύνη
AR (Augmented Reality)	Επαυξημένη Πραγματικότητα
Big Data	Μεγάλα Δεδομένα
CIDOC-CRM	Πρότυπο CIDOC-CRM
Collection Management System (CMS)	Σύστημα Διαχείρισης Συλλογών
Conservation Report	Έκθεση Συντήρησης
Data Governance	Διακυβέρνηση Δεδομένων
Data Interoperability	Διαλειτουργικότητα Δεδομένων
Data silos	Σιλό Δεδομένων
Digital Asset Management (DAM)	Διαχείριση Ψηφιακών Πόρων
Digital Twin	Ψηφιακό Δίδυμο
GDPR (General Data Protection Regulation)	Γενικός Κανονισμός Προστασίας Δεδομένων
Knowledge Graphs	Γραφήματα Γνώσης
Linked Open Data (LOD)	Διασυνδεδεμένα Ανοιχτά Δεδομένα

Metadata	Μεταδεδομένα
Open Access	Ανοικτή Πρόσβαση
Provenance	Προέλευση / Ιστορικό Κατοχής
Semantic Markup	Σημασιολογική Σήμανση
User-Generated Content (UGC)	Περιεχόμενο που Δημιουργείται από Χρήστες
Visitor Analytics	Ανάλυση Επισκεπτών
VR (Virtual Reality)	Εικονική Πραγματικότητα