Huge Content Data Management for Numerous Applications in the Museum

Jae-Ho Lee, Chan-Woo Park, and Hee-Kwon Kim
Creative Content Research Division, Electronics and Telecommunications Research Institute, Daejeon, Korea
jhlee3@etri.re.kr, gamer@etri.re.kr, hkkim79@etri.re.kr

2024 Fifteenth International Conference on Ubiquitous and Future Networks (ICUFN) | 979-8-3503-8529-8/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICUFN61752.2024.10625592

***Abstract*—The digital transformation wave is significantly reshaping museum operations and amplifying the demand for robust data management technologies. As museums grapple with vast amounts of data, efficient solutions are essential to harness cultural heritage information effectively. This paper explores the critical role of data management technology in supporting realistic content creation, preservation, management, and educational endeavors within the museum context. Standardizing cultural heritage digital data, generated across diverse museum departments, emerges as a pivotal step toward successful digital transformation. Our ongoing research centers on developing an AI-based digital heritage platform tailored to traditional cultural artifacts. By establishing comprehensive standards for data creation, transformation, processing, analysis, and visualization, we aim to efficiently manage the substantial data volumes generated by museums.**

***Index Terms*—Huge Content, Digital Heritage, Cultural Heritage, Intelligent Platform, Digital Transformation**

I. Introduction

The advent of digital technologies has revolutionized the museum landscape, prompting institutions to adapt swiftly to the digital era. Museums now face the dual challenge of preserving cultural heritage while embracing innovative approaches to engage audiences. Central to this transformation is the effective management of cultural heritage data, which encompasses artifacts, historical records, multimedia content, and scholarly research. This digital shift enhances the visitor experience with interactive exhibits and virtual tours, while also presenting new challenges in data preservation and curation. Museums must now develop robust digital strategies to effectively store, catalog, and share their collections with a global audience [1]. As a result, there is an increasing need to develop a digital heritage-sharing platform for traditional cultural heritage content. Such platforms enhance accessibility to cultural heritage, offering opportunities to introduce traditional content in digital form to a broader audience. Additionally, they can be utilized for educational purposes, presenting ways to experience cultural heritage from a contemporary perspective [2] [3] [4] [5]. In this paper, we delve into the multifaceted role of data management technology within museums. We emphasize the need for tailored solutions that accommodate the unique characteristics of cultural heritage data. Furthermore, we advocate for standardized practices across museum departments to facilitate seamless collaboration and data exchange. Our research focuses on the development of an AI-driven digital heritage platform, designed to enhance access, preservation, and educational outreach. The below describes challenges in cultural heritage data management.

- (Data Volume and Diversity) Cultural heritage data spans a wide spectrum, from digitized manuscripts and archaeological artifacts to immersive virtual reality experiences. The sheer volume of data generated by museums necessitates efficient storage, retrieval, and dissemination mechanisms. Moreover, the diversity of data formats—ranging from high-resolution images to 3D models—requires adaptable solutions.
- (Realistic Content Creation) Museums increasingly leverage digital technologies to create immersive and realistic content. Virtual reconstructions, interactive exhibits, and augmented reality experiences enhance visitor engagement. Effective data management ensures that these creations remain accurate, contextually rich, and accessible.
- (Preservation Challenges) Preserving cultural heritage data involves safeguarding it against degradation, loss, and obsolescence. Long-term storage, metadata preservation, and data migration strategies are critical components of any preservation framework. Additionally, managing rights and permissions ensures ethical and legal compliance.
- (Educational Use and Outreach) Museums serve as educational hubs, disseminating knowledge to diverse audiences. Cultural heritage data plays a pivotal role in educational programs, research collaborations, and public engagement. Robust data management facilitates seamless sharing, enabling educators, scholars, and students to explore and analyze artifacts remotely.
- (Policy perspective) The Korea government has announced a strategy for using artificial intelligence to digitize cultural assets and create digital intellectual property. The plan includes developing a system for converting existing digital and analog data into knowledge-based intellectual property data, using the national integrated platform and digital government big data. This announcement was made in February 2021.

Towards an AI-Based digital heritage platform, our ongoing research centers on developing an AI-driven digital heritage platform. Key objectives include:

1) Standardization: Establishing uniform data standards across museum departments ensures interoperability and efficient data exchange. We propose guidelines for meta-

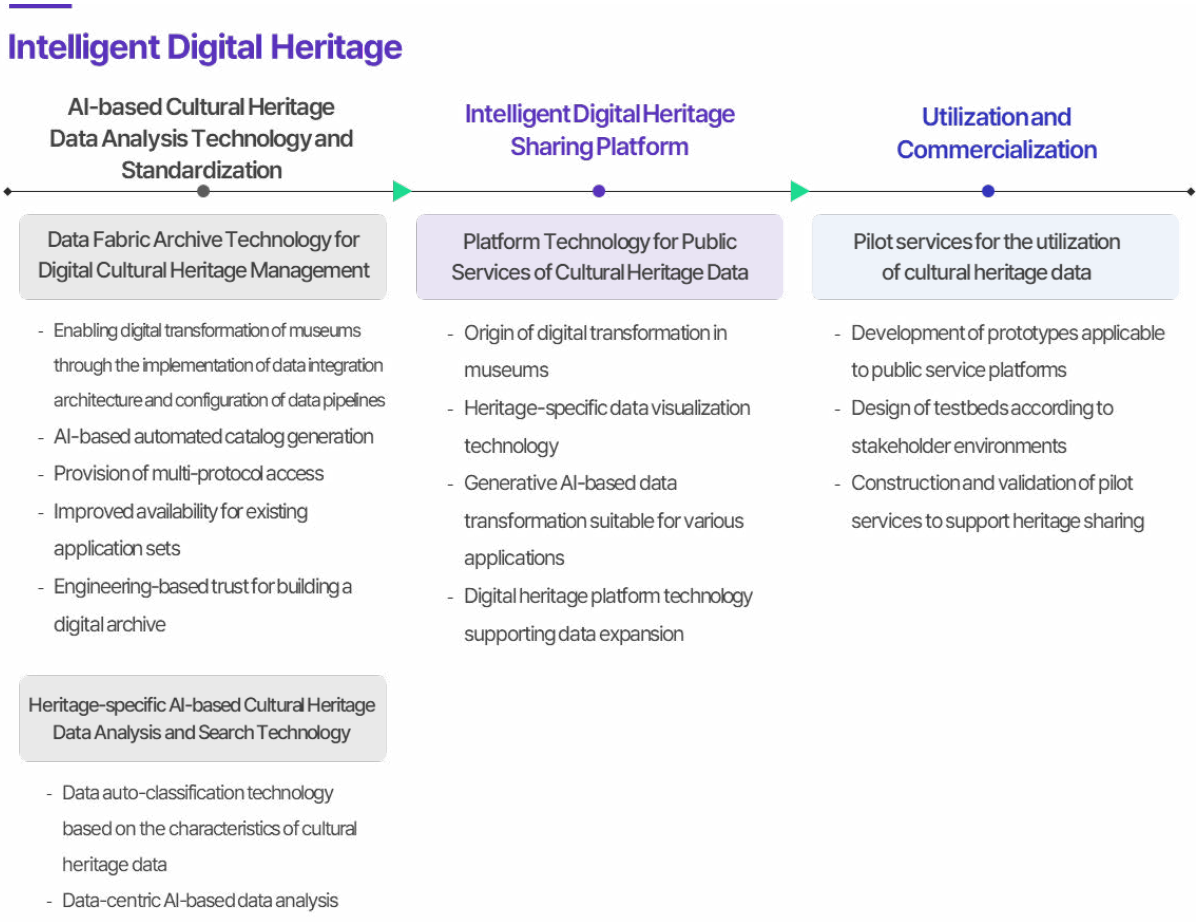

Fig. 1. Research and development items of intelligent digital heritage sharing platform.

data schemas, data formats, and ontologies.

2) Semantic Enrichment: Leveraging AI techniques, we enrich cultural heritage data with contextual information. Natural language processing, image recognition, and linked data enhance data discoverability and interpretation.
3) Dynamic Visualization: Our platform incorporates advanced visualization tools, allowing users to explore 3D models, historical timelines, and spatial reconstructions. Real-time rendering and interactive interfaces enhance the visitor experience.
4) Collaborative Workflows: By integrating collaborative workflows, our platform fosters interdisciplinary research. Curators, historians, conservators, and technologists collaborate seamlessly, contributing to holistic data management.

II. Multi-purpose Universal Cultural Heritage (MUCH)

The project, as outlined in Figure 1, aims to develop technology that can efficiently manage and utilize traditional cultural heritage content data held by museums nationwide. To achieve this, the project will concentrate on two pivotal areas: firstly, the development of a data fabric archive technology to efficiently store and retrieve cultural heritage content; and secondly, the creation of an AI-based technology for the analysis and search of cultural heritage data. These technologies will standardize and convert existing cultural heritage digital data, thereby enhancing accessibility and usability for a wider audience. Furthermore, the project endeavors to develop a cultural heritage digital standard-sharing platform, which will serve as a hub for accessing and disseminating standardized cultural heritage data. The Data-Fabric archive technology will be instrumental in this initiative, tasked with analyzing and managing high-quality, standardized cultural heritage digital data. These technological advancements are not only necessary but also serve as the cornerstone for the future of cultural heritage content management and public engagement.

Numerous studies are leveraging data to develop AI technology aimed at automatically classifying cultural heritage data based on inherent characteristics [6]. In addition to classification, efforts are underway to develop technologies that enable advanced search capabilities and personalized recommendations through catalog tagging of cultural heritage data. The primary objectives of these research and development efforts are to create generative AI-based technology for cultural heritage data conversion and to facilitate the authoring and

publishing of realistic content, utilizing automatic conversion technology designed for the preservation or research of data [7] [8] [9] [10]. The subsequent subsections of this chapter will detail the methodologies and outcomes of our research and development initiatives.

This study aims to present the comprehensive process of constructing a digital platform for cultural heritage, along with the requisite technological components. Specifically, we will offer a detailed exposition of the AI technologies employed, data structures, and associated techniques, including quantitative comparisons, once the technical development reaches an advanced stage. The central objective of this paper is to outline the fundamental elements necessary for establishing a cultural heritage digital platform using AI technologies and to introduce the ongoing efforts in this domain.

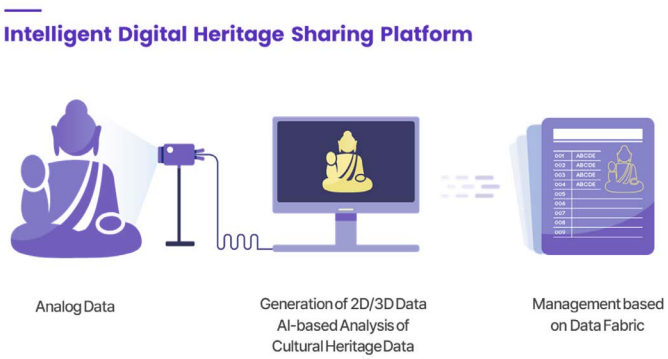

Fig. 2. Intelligent Cultural Heritage Sharing platform.

A. *Data fabric-based archive design*

A design for storing cultural heritage data types was created by selecting representative structured and unstructured data from museum cultural heritage and classifying them. The National Museum of Korea has collected over 2 million pieces of unstructured and unrefined digital data from every department to date. The metadata was designed by comparing and analyzing the data values of the standard artifact management system with the e-museum, which was the target of data construction. This study represents Korea's first integrated metadata-based archiving data type and storage structure technology study for digital heritage. The goal is to establish a 3D data archive standard specialized for digital heritage and to conduct data classification and relationship analysis, focusing on 3D assets of artifacts owned by the National Museum of Korea.

A recent study analyzed essential storage data for structural data by comparing cultural heritage data related to research, education, and content from seven different departments, including the Department of Conservation Science and the Department of Antiquities of the National Museum of Korea. The study compared the quality of the initial generation method of existing digital data with the latest digital data, and the results were used to update the 3D digital data creation guidelines. The aim is to maintain consistency and improve efficiency in data generation.

The digital data stored in the central museum was unstructured and unrefined. To comply with the W3C Semantic Web standard, the data was analyzed, and a configuration was designed that encompasses the existing digital data. This included performing database configuration design for the conversion and management of existing digital data and creating guidelines for Giga Pixel, 3D, RTI, CT, Photogrammetry. A metadata design was also developed, which includes 3D data archiving standards for heterogeneous objects specific to digital heritage. By redefining how to utilize files related to unstructured and unrefined 3D data files, the basic data structure for 3D data management was created. By analyzing the format of 3D model data, items can be determined to standardize the types and functions of files used in 3D Viewer. This aids in developing integration and conversion methods for 3D data construction technology based on the intended purpose of use.

Currently, numerous studies are being conducted to develop data processing technology specializing in the artificial intelligence analysis of cultural heritage [11] [12] [13]. Our aim is to establish automatic classification and tagging of cultural heritage on AI-based digital platforms by defining relationships between them, as illustrated in figure 2. To achieve this, the current DAMS(Data Archive Management System) data from the National Museum of Korea is analyzed to identify AI-based automatic relationships between cultural heritage items. An additional survey is conducted to determine cultural property classification, focusing on traditional systems such as the Korean Decimal Classification (KDC), the Cultural Heritage Administration's National Cultural Heritage Portal classification, the National Museum of Korea's e-Museum cultural property classification, and Korean local culture.

Furthermore, extensive research is conducted to classify cultural properties and define metadata. This includes analyzing exhibition directory classification, planning research materials for the production and distribution of cultural property 3D assets, using the National Museum Standard Artifact Management System, and managing folk record data. It also involves identifying cultural heritage tagging/relationships and specifying context-based metadata elements and formats for the cultural heritage database. To develop asset-based intelligent curation and service operation technology for realistic cultural heritage experiences, CIDOC CRM and CIDOC CRM Extension are referenced and compared in our research.

Our analysis identified limitations in current approaches for storing 3D cultural heritage data generated or captured through novel technologies. Unlike traditional 3D data, newly created cultural heritage data comprises a multitude of object-level data units. Consequently, a well-defined framework was essential to represent this data effectively. Furthermore, establishing structural relationships was crucial to connect digital information generated during preservation, processing, and exhibition. In our research, we addressed these challenges by augmenting existing methodologies.

B. Cultural heritage data analysis technology

A new technology has been developed to classify different types of cultural heritage data using AI. This technology considers the structure of cultural heritage data and employs a digital cultural heritage ontology for classification. The ontology was developed by analyzing traditional cultural heritage classification systems, research methodologies, museum metadata creation manuals, and existing examples of tagging and relationship definition. To create an efficient metadata structure utilizing the relationships between different data types, the creators visited the National Museum of Korea to analyze the data's status and characteristics. They designed an ontology structure that visualizes connections between data and relationships between cultural heritage items.

The concept of digital data is extended in the standard artifact management system to capture the essential metadata of heterogeneous data. This was achieved by considering the requirements of the National Museum of Korea and studying existing metadata design cases. The focus was on creating an efficient ontology using hierarchical classification criteria, which allows for a more nuanced approach than existing type classification criteria. The definition of a linked data layer and the structural relationship between related data were also considered.

The data catalog includes various attributes such as collection identification code, cultural property name, quantity, nationality, material, use/function classification, thematic classification, excavation site, acquisition information, pattern decoration, excavation information, insurance information, and copyright. These attributes are arranged hierarchically, with operations at the highest level of abstraction representing the conceptual essence of the resources listed, such as author, language, and subject. A single work can have multiple physical representations, each with specific information like publisher, location, date of publication, and format. The individual data, also known as an item, is a substructure of an instance and refers to the smallest unit that can be used as an independent object. A large dataset was created for AI analysis by defining data attributes.

The dataset was categorized into 10 major categories and 92 subcategories. The study focused on the automatic recognition and extraction technology of object names found in traditional culture museums using deep learning technology [14] [15]. As a result, an accuracy rate of 89% was achieved in the automatic tagging of cultural heritage.

C. Standard classification technology for cultural heritage digital data

Museums collect a vast array of cultural heritage data, encompassing both structured and unstructured forms. Developing acceptable standardization processes is crucial for the efficient management of new digital data, including unrefined digital files. Our research has led to the proposal of a standardized method for data generation and construction, informed by the asset construction performance at the National Museum of Korea. This method encompasses considerations for technology, equipment selection, environment construction, safety, and precautions, all tailored to the unique characteristics of cultural heritage.

Post-processing techniques are employed to transform general sources, whether image files or 3D numerical information, into ultra-high resolution images, reflectance change images, polygons, and texture mapping sources. We propose new standards for the creation and construction of data involving professional creative work, such as 3D, 2D, RTI, and photogrammetry, tailored to the specific type of digital cultural heritage data.

A novel data structure is proposed to link various types of digital cultural heritage data, comprising three levels: Work, Instance, and Item. Additionally, a Project Level is introduced to accommodate the digital data needs of the National Museum of Korea. To bridge the gap between existing and new data, we have developed a method for cross-referencing through the standardization of a knowledge graph-based digital cultural heritage data generation.

Crucial metadata is defined to articulate the relationship between digital cultural heritage data and other usable data, utilizing an open knowledge graph digital cultural heritage data ontology. Our ontology is designed based on a standard generation model, and the pilot model follows the characteristics of the BIBFRAME model.

D. Public service cultural heritage-based framework

A knowledge graph data conversion module has been developed to facilitate the design of an open digital cultural heritage platform based on knowledge graphs. The module employs a mapping table that aligns with the columns of the original data to ensure the converted data adheres to the original data standards defined by W3C. A code generation process has been established to transform the data into triple data format, utilizing mapping rules for source data columns based on relationship properties such as Object and DataType.

A Java language-based triple data transformation module was developed, which complies with the established mapping rules and accommodates the source data type. The generated triple data is published on the web, and a function has been defined that allows users to navigate to other data linked via hyperlinks on the publication page. Additionally, a file format storage function has been developed to support triple data formats, including Turtle, RDF/XML, JSON/LD, and N3.

A pilot service has been launched to visualize archives, accessible through a public service platform as depicted in Figure 3. Our objective is to specialize in cultural heritage visualization by employing various methods that effectively convey the information encapsulated within digital cultural heritage data structured in an ontology [16] [17] [18]. The pilot visualization service utilizes metadata based on layers such as Work, Instance, Item, and Project, which are structured according to the standard classification of cultural property digital data.

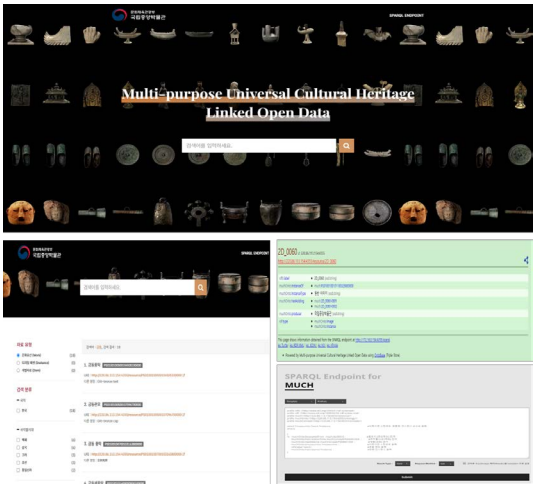

Fig. 3. Multi-purpose Universal Cultural Heritage (MUCH) platform test homepage.

We have tested five visualization models—list, grid, network, cloud, and timeline—and produced pilot content that has undergone user satisfaction testing and demonstrations. A satisfaction survey was carried out for our archive-based data visualization prototype, which was exhibited at the 2023 World National Heritage Industry Expo, to gauge participant satisfaction with the data visualization.

A framework has been proposed for public service heritage-based data that aligns the columns of the original data with W3C standards, incorporating support for SPARQL(Simple Protocol and RDF Query Language), triple data management, and leveraging knowledge graph-based open data platform technologies. We have also developed a pilot service for a public service heritage-based data platform that adheres to W3C standards, achieving an 85

As part of our efforts to develop a technology platform for cultural heritage, we created a specialized algorithm using Style Transfer for AI analysis. This algorithm employs a CNN to separate and recombine image content and style, utilizing the VGG19 network's layers to extract information about different image aspects. An image conversion network was developed using a perceptual loss function, commonly employed to digitally restore damaged cultural heritage data using an inpainting algorithm. Inpainting utilizes a pre-trained Denoising Diffusion Probabilistic Model to fill in missing image parts based on known regions, employing a 'resampling' method to enhance the accuracy of the generated images. Our results were obtained by processing 160,000 eras of image data collected from 1,808,321 cultural heritage images. The evaluation metrics used were IS (2.79), FID (28.23) and MS-SSIM (0.76).

E. Archive-based data visualization service

A pilot service was developed to visualize cultural heritage-specific data archives that reflect the requirements of the National Museum of Korea. The service is based on a WebGL-based 3D model viewer and is platform-independent. The main purpose behind developing this service is to provide core platform functions and UI/UX so that museum staff can utilize public services required for cultural heritage research or content planning. Additionally, an RTI (Reflectance Transformation Imaging) viewer was developed for high-precision analysis of cultural heritage. To ensure effective viewer functionality, image tiling technology was applied to the PTM format obtained using the PTM to RTI Image Converter file converter. By adding the image tiling viewer function to the online web viewer, the RTI viewer function for cultural heritage videos has been implemented in the same online environment.

The construction of Heritage PBR (Physically Based Rendering) archive data aimed to preserve cultural heritage digital assets and enable ultra-high-quality visualization services. Each artifact's characteristics were considered when preparing 2D and 3D assets. A data acquisition plan was established for each artifact to ensure accurate processing. High-precision scanners were used for small artifacts (5-30 cm), while medium-precision and high-precision scanners were used for medium-sized artifacts (50 cm or more), depending on their surface complexity and shape. Template data was created to simulate values such as reflection, transparency, blur, and projection according to the material and light source environment during shooting, making it easier for users to utilize the data.

The Archive-based Data Visualization Service is a platform designed to provide secure storage and ultra-high-quality visualization services for digital data acquired from cultural heritage artifacts. The service has obtained 3D assets (50 artifacts, 60 points), RTI (10 artifacts, 15 points), and Gigapixel (5 artifacts, 5 points) data for a total of 56 artifacts (66 points) from several museums, including the National Museum of Korea. Around 30 of these cultural assets are scheduled to be used in the National Museum of Korea's Goguryeo monument project, which is set to open in January 2024.

Additionally, a 3D model viewer that uses WebGL technology to meet the needs of various organizations has been developed [19]. This viewer allows users to explore and interact with 3D models through a web interface. To ensure high-quality 3D data viewing in various online environments, the user experience has been optimized by compressing the glTF format to the Nexus format. The RTI viewer is responsible for rendering complex images based on WebGL and implements the overall viewer functionality.

Furthermore, the viewer has constructed physically based rendering (PBR) materials for 34 types of cultural heritage materials, such as metals, ceramics, glass, textiles, and leather. Using the Cinema 4D program, the characteristics of the PBR material were applied, and position and intensity changes of lighting were checked to create PBR data that closely resembles reality, as shown in Figure 4.

III. Discussions and Conclusions

As museums embrace digital transformation, data management technology emerges as a linchpin. Our research endeavors to bridge the gap between cultural heritage and cutting-

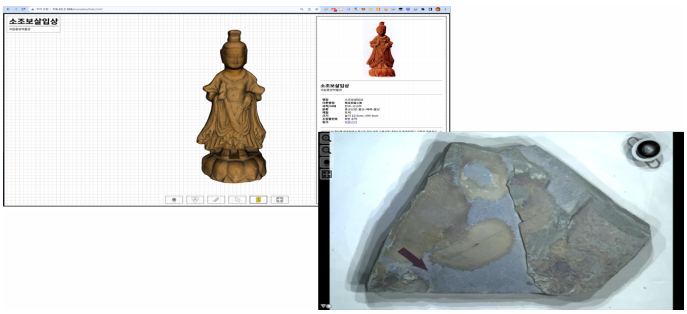

Fig. 4. Asset Viewer and RTI viewer in Multi-purpose Universal Cultural Heritage (MUCH) platform.

edge AI solutions, ensuring that museums remain vibrant repositories of knowledge accessible to global audiences.

Currently, our focus lies in developing generative AI technology that can effectively utilize cultural heritage data across multiple devices, including smartphones, web platforms, and head-mounted displays (HMDs). This technology encompasses the creation of a cultural heritage data catalog with tagging-based search and recommendation systems, as well as the development of various data acquisition and classification techniques. Our primary emphasis is on Data-Centric AI-based classification of cultural heritage data types.

Furthermore, we plan to conduct research on NeRF (Neural Radiance Fields)-based 2D-3D data generation technology, aligning it with digital cultural heritage standardization efforts. Our ultimate goal is to define semantic relationships between data elements by designing them as an ontology. For ontology design, we will specify classes, instances, relationships, and properties relevant to various cultural heritage data domains.

This project is a direct response to the policy demands of the National Museum of Korea. We have engaged in direct discussions with various stakeholders, including curators, system managers, exhibition planners, and designers, to gather user requirements and address specific issues. Our aim is to enhance the usability for curators, who are the primary users. We base our research on the requirements expressed by related departments within the National Museum and incorporate feedback from the pilot platform into technology development requests and design guidelines. The launch briefing held at the National Museum of Korea reaffirmed the necessity and vision of this project, and ongoing interactions with museum directors and curators from relevant departments will guide the continuous development process.

As we move forward, we plan to discuss our development roadmap and determine when to apply platform-specific technologies required by relevant departments. This strategic alignment is crucial for effective system utilization. In today's digital era, recognizing the importance of handling unrefined and unstructured data smoothly, we aim to establish a systematic data management system. This system is expected to facilitate the recycling and organized management of diverse data generated through existing research and preservation efforts. Our goal is to enable users to quickly and accurately access the data they need from the vast pool of existing digital resources.

This paper synthesizes the critical aspects of cultural heritage data management, emphasizing the intersection of technology, preservation, and education. By framing our ongoing research within the context of an AI-based digital heritage platform, we contribute to the evolving landscape of museum practices. The journey toward efficient data management continues, guided by the rich legacy of cultural artifacts and the promise of technological innovation.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency (Project Number: RS-2023-00219579)

REFERENCES

[1] E.Jang, J. Lee, G. Shin, J. Choi, and Y. Jo, "Coexistence and Conservation Building an Intelligent Curation Platform," National Museum of Korea, pp. 28-29, September 2022.
[2] https://www.emuseum.go.kr/main.
[3] www.wikiart.org.
[4] www.europeana.eu.
[5] www.museumnext.com.
[6] G. Srezoski, et al., "Omniart: Multi-task Deep Learning for Artistic Data Analysis," arXiv preprint arXiv:1708.00684, CoRR, Aug. 2017.
[7] DeOldify: A Deep Learning based project for colorizing and restoring old images and videos, Jason Antic, 2018.
[8] J. Kim, J.k. Lee, and K.M. Lee, "Accurate image super-resolution using very deep convolutional networks," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654, 2016.
[9] A.Gunawan, S. Kim, H. Sim, J. Lee, and M. Kim, "Modernizing Old Photos Using Multiple References via Photorealistic Style Transfer," CVPR, pp. 12460-12469, 2023.
[10] C. Park, H. Kim, and J. Lee, "A Study on the Restoration Method of Black and White Image Cultural Heritage Using Deep Learning Based on Open Source," Summer Annual Conference of IEIE, pp. 2139-2142, June 2023.
[11] J. Lee, H. Lee, H. Kim, and C. Park, "Implementation of Digital Transformation for Korean Traditional Heritage," FCV, pp. 257-270. 2022.
[12] J. Lee, H. Lee, H. Kim, and C. Park "Studies on intelligent curation for the Korean traditional cultural heritage,"ICAIIC 2022.
[13] J. Lee, H. Lee, H. Kim, and C. Park "CHIC: cultural heritage intelligent curation platform," ICEIC 2021.
[14] B. Nie, R. Ding, P. Xie, F. Huang, C., Qian, and L. Si, "Knowledge aware named entity recognition with alleviating heterogeneity," AAAI Conf. Artificial Intelligent, vol 35, no. 15, pp. 13595–13603, 2021.
[15] Q. Tran, A. MacKinlay, and A.j. Yepes, "Named entity recognition with stack residual ISTM and trainable bias decoding. arXiv:1706.07598.
[16] J. Lee, H. Shin, and S. Im, "Development of Data Labelling and Visualization Technologies for Cultural Heritage Intelligent Curation," Heritage 6, no. 3, pp. 2735-2751, 2023.
[17] F. Windhager, et al., "Visualization of Cultural Heritage Collection Data: State of the Art and Future Challenges," IEEE Transactions on Visualization and Computer Graphics, 2018, Vol. 25, No. 6, pp. 2311-2330.
[18] B. Kim, and J. Lee, "Analysis of Cases and Proposal of a Tool to Visualize the Relationship Information of Cultural Heritage," KoCon 2022, pp. 279-280, July 2022.
[19] C. Park, H. Kim, and J. Lee, "Historical Asset Viewer for Efficient Management of Cultural Heritage," ICCE-asia 2023.