Workflows for analysing and utilizing large-scale 3D datasets of cultural heritage

Sander Münster [1], Paweł Marciniak [2], Mikołaj Węgrzynowski [2], Francisco de Arriba Pérez [3], Silvia García Méndez [3], Dominik Ukolov [1], Simone Rigon [4], Elisa Mariarosaria Farella [4], Fabio Remondino [4], Ying Sun [1], Alexandru Stan [5]

[1] Digital Humanities, Friedrich-Schiller-Universität Jena, Germany – (sander.muenster, dominik.ukolov, ying.sun)@uni-jena.de
[2] Poznań Supercomputing and Networking Center, Poland – (pawelmar, mwegrzyn)@man.poznan.pl
[3] University of Vigo, Spain – (farriba, sgarcia)@gti.uvigo.es
[4] 3D Optical Metrology (3DOM) unit, Bruno Kessler Foundation (FBK), Trento, Italy – (srigon, elifarella, remondino)@fbk.eu
[5] IN2 Digital Innovations GmbH, Lindau, Germany – as@in-two.com

Keywords: 3D models, Data Spaces, Museum Applications, Education, Cultural Tourism.

Abstract

For 3D content, large-scale datasets, such as Objaverse or ShapeNet, and repositories, such as Sketchfab, have been compiled. Within the European 3DBigDataSpace project, a consortium of 10 partners assesses open licensed 3D models to select and retrieve those models particularly representing cultural heritage objects in Europe to aggregate them into the European Data Space. As key part of that work is the classification and geolocalization of 3D content, with mesh models viewable via different viewers and tested in different scenarios such as museum exhibitions, cultural tourism, or education. This article highlights the steps taken (1) to compile a large-scale pool of 3D assets of cultural heritage and ready-to-use viewer applications and (2) to enrich and (3) utilize them in various settings.

1. INTRODUCTION

Current challenges with using 3D models in museums, education or tourism include the high levels of customization needed and the limited availability of cultural heritage 3D models (Münster, 2023). In this article, we present the first steps taken (1) to compile a large-scale pool of 3D assets of cultural heritage and ready-to-use viewer applications and (2) to enrich and (3) to employ them in various settings to enhance and validate the usability of 3D models. While the data collection and viewer development have already reached an intermediate stage, the validation within use cases is currently ongoing.

2. STATE OF THE ART

Datasets and collections
For the 3D content, large-scale datasets such as Objaverse including 10.2 million 3D models (Deitke et al., 2023) or ShapeNet counting 50k 3D assets (Chang et al., 2015), as well as repositories like Sketchfab hosting several 100,000 heritage items (Flynn, 2019) have been compiled and shared. Large-scale 2D/3D datasets include MVImgNet2 (Wu et al., 2024) and MegaScenes (Tung et al., 2024). In addition, campaigns such as Scan the World[1] or Global Digital Heritage[2] make a collective effort to digitize cultural heritage on a large scale. Various researchers investigated the retrieval of sufficient metadata from 3D objects and metadata to spatialize and temporalize this material (e.g. (Orzechowski et al., 2025; Münster, 2023)).

Metadata enrichment
Geolocalization: Geo-based data is visualized in various web-based portals, increasingly taking a 3D approach, with 4Dcities (Schindler and Dellaert, 2012) being one of the first to make historical photographs accessible on the web in relation to 3D models. Other applications introduced time-evolving 3D city models incorporating multimedia such as photographs and other historical documents (Blettery et al., 2021; Jaillot, 2020). However, precise spatial and temporal location of the data remains challenging, requiring either a rich set of metadata or manual work. Several rephotography approaches and applications support the spatialization of photographs, but still require manual input (Schaffland et al., 2020). Other approaches include the geolocalization via images (Wang et al., 2024; Pramanick et al., 2022; Cepeda et al., 2023; Kulkarni et al., 2024; Xu et al., 2024) and texts, e.g. (Singh and Aneja, 2024). For imagery, several methods exist for homogeneous image blocks (Sattler et al., 2018; Sarlin et al., 2019). The problem becomes increasingly complex for varying radiometric and geometric conditions, especially relevant for historical photographs (Maiwald, 2022). More recent geolocalization frameworks integrate multiple stages of processing – e.g. by style similarity, identification of captions, or geometric reconstruction. Examples are Plonk (Dufour et al., 2024), Pigeotto (Haas et al., 2024), GeoCLIP (Vivanco et al., 2023) and OrienterNet (Sarlin et al., 2023).

Object classification: Efficient retrieval and exploration of historical images are based on visual similarity and content-based features. However, traditional machine learning technologies currently require large-scale training data (Fiorucci et al., 2020; Radovic et al., 2017; Aiger et al., 2017; Münster et al., 2024b) and are primarily capable of recognizing well-documented and visually distinctive landmark buildings (Münster et al., 2021) but fail to deal with less distinctive architecture, such as houses of similar style. Even using more advanced approaches or combining different algorithms (Maiwald et al., 2021) only enables the realization of prototypic scenarios (Gominski et al., 2019; Morelli et al., 2022). Since these approaches are mostly trained on photographs, various projects are using training material from 3D datasets such as Objaverse to train classification approaches (Lin et al., 2025).

Data processing: Another challenge relates to processing recorded 3D cultural heritage datasets. Various commercial companies provide tool ecosystems which include format conversions, metadata assignment, and the creation of browser-based 3D views. The best known examples include Sketchfab for 3D data and Luma.AI for Gaussian Splats.[3] Several European and non-commercial services such as Share3D,[4] the Eureka3D Data Hub,[5]

[1] https://www.myminifactory.com/scantheworld/
[2] https://globaldigitalheritage.org/
[3] https://lumalabs.ai/interactive-scenes
[4] https://www.share3d.eu/en/
[5] https://eureka3d.eu/eureka3d-data-hub/

and the Zenodo toolbox (Münster et al., 2024a) provide pipelines for collecting and aggregating 3D datasets in Europeana. Since those tools come with a predefined pipeline, frameworks like Smithsonian Cook[6] provide means to integrate multiple tools into versatile pipelines. This approach enables researchers to setup flexible pipelines but also to integrate services distributed by other parties or developers and has therefore been chosen by our project for the development of a Heritage Data Processor framework.

3D Viewer

Various viewer stacks are available to visualize 3D cultural heritage objects (Overviews: (Fung et al., 2021; Bajena et al., 2022; European Commission, 2022; Storeide et al., 2023; Münster, 2023). To integrate multiple viewer stacks supporting specific purposes and communities, the German DFG 3D Viewer project utilizes a flexible wrapper structure that can accommodate various viewers and has been utilized and extended within the 3DBigDataSpace project (Münster et al., in press).

3. DATA COLLECTION

The initial dataset utilized in 3DBigDataSpace stems from different data collections and was compiled between 11/2023 and 4/2025. To retrieve legally accessible content, we selected CC-0 or CC-BY licensed content only. For data retrieval, we used a series of server-side scripts in Python and PHP feeding into an SQL database and Unix file storage. From the pool of 11.4 M metadata sets currently collected and partly processed, a subset of ca. 130,000 3D modelshave is utilized for the following study (Table 1).

Data source	**No. of 3D objects**	**Description**
Europeana	8,708	The Europeana 3D dataset contains validated metadata and is utilized to provide Ground Truth data. It was retrieved via the Europeana Python Framework.[7]
Objaverse 1.0	55,614	The Objaverse 1.0 dataset includes 800,000 3D objects with 55,000 datasets classified as Cultural Heritage. It was compiled by the Paul Allen institute. The datasets are mainly retrieved from open licensed content held by Sketchfab.
5DCulture	8,406	The 5DCulture dataset was compiled in the project of same name in 2024. The dataset includes various mainly low-poly models of single buildings from different age in the cities of Trento, Sion, Amsterdam, Dresden, and Jena. The dataset was used to test the Zenodo pipeline.
Smithsonian	3,685	A set of openly licensed 3D models from the Smithsonian collection that was processed in mid-2025.
Scan the World	12,448	An ecosystem to freely share digital, 3D scanned cultural artefacts for physical 3D printing, processed in 10/2025.
Megascenes	69,475	Public image data showing architecture and immovable artworks, partially reconstructed as sparse point clouds reconstructed. From the full dataset 3D models with sparse point clouds larger than 1kb were selected for the 3D dataset.
LiDAR 3D Buildings	1, 374	LiDAR datasets segmented via OSM ground plots into multiresolution single building models. These models are used for mapping and multi-level-of-detail approaches (Münster et al., 2025b).

Table 1. Processed datasets.

4. End-to-end integration

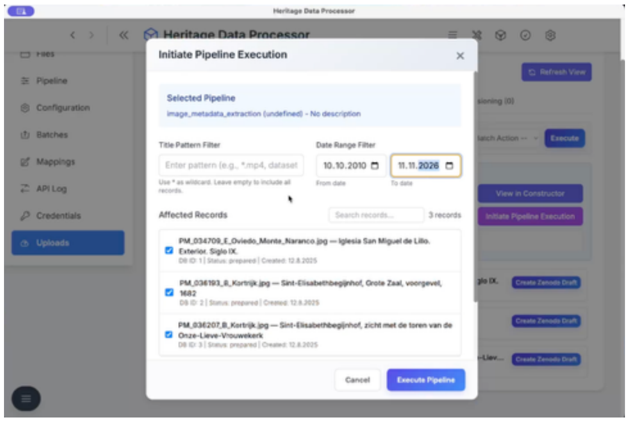

Figure 1. Screenshot of the Heritage Data Processor interface.

The Heritage Data Processor (HDP) is a GUI, API, and Command-Line-driven application designed for the comprehensive processing, enrichment, and management of digital assets, with a primary focus on multimodality (Figure 1).[8] It was developed within the 3DBigDataSpace project to provide a framework for integrating services provided by the different partners and to set up processing pipelines. The processor leverages a combination of machine learning models, large language models (LLMs), and external APIs to automate complex data processing workflows. The HDP is intended to be used for the construction of complex pipelines, based on a modular system of pipeline components, with stable storage as its endpoint using Zenodo. Together with the predecessor Zenodo toolbox, around 65,000 3D mesh models and another 65,000 photographs has been processed and stored in Zenodo (Münster et al., 2024a).

5. 3D MODEL OPTIMIZATION

A MeshOptimizer tool was developed within the project to facilitate large-scale 3D mesh processing and improve model visualization across various platforms. The tool is a fully automated and modular 3D mesh processing pipeline, designed for optimizing both geometry and visual texture quality. The framework supports both decimation and remeshing, as well as texture improvement through the generation of Physically Based Rendering (PBR) maps, including diffuse, ambient occlusion, roughness, and normal maps. Geometry optimization is designed to balance mesh simplification and object fidelity. The framework, containerized via Docker and implemented in Python and C++, leverages Blender for texture processing and geometry decimation, PartUV for efficient UV mapping,[9] and CGAL's adaptive isotropic remeshing C++ libraries for remeshing, surface analysis, and geometric transformations. The Hausdorff distance is iteratively computed to analyse the geometric deviation at each remeshing stage, and ensures adaptive optimization while preserving the integrity of the original model (see Figure 2).

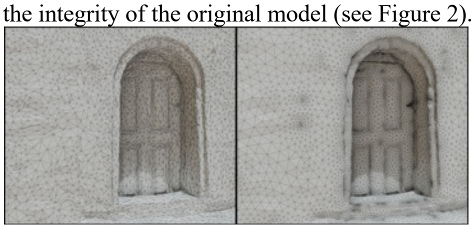

Figure 2. An example of mesh simplification obtained with the MeshOptimizer tool – original (left) and optimized (right).

[6] https://smithsonian.github.io/dpo-cook/
[7] https://github.com/europeana/rd-europeana-python-api
[8] https://github.com/Digital-Humanities-Jena/heritage-data-processor
[9] https://github.com/EricWang12/PartUV

6. DATA AND METADATA ENRICHMENT

Data and metadata processing includes (a) data processing to pre-process captions and render images of the models; based on which the data is (b) categorized and (c) geolocalized. For all those steps, we currently employ different models (Table 2). A main rationale for selecting models has been to assess the trade-off between quality and computing costs.

Text preprocessing	• Language detection: Fasttext[10] • Language translation: nllb-distilled-1.3B[11] • Compiling short descriptions from text: distilbart-cnn-12-6[12] • Descriptions from images: LLAVAnext 1.6 / Mistral 7B[13]
Image rendering	• Rendering: Blender for Python: Fastrenderer and Cycles
Text-based categorization	• deberta-v3-large-zeroshot-v2.0[14] • QWEN 3.0-14B MLX[15] • QWEN 3.0-32B MLX[16]
Image-based classification	• LLAVAnext / Mistral 7B • CLIP / ViT-B/32[17] • Qwen3-VL-30B-A3B-Instruct (4-bit quantized) • VGG-16
Text-based geolocalization	• LLAMA 3.2[18] • Distilled Deepseek R-1 + QWEN 7B[19] • QWEN 3.0-14B MLX • QWEN 3.0-32B MLX • Deepseek R-1 SaaS[20] • Spacy with core_en_web-lg model[21] • Multi-label classifier: SciKitLearn[22]
Image-based geolocalization	• Plonk[23] • OrienterNet[24] • GeoCLIP[25] • Geocoding: Geopy with OSM / Nomatim and GooleV3 API for resolving coordinates

Table 2. Models currently used for pipeline integration.

Text preprocessing
For around 1/3 of the 3D datasets, titles and descriptions are available only in other languages than English (Figure 3).

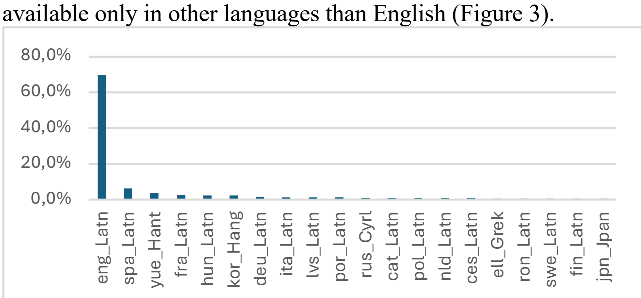

Figure 3. Input language text classification, TOP 20, classified via a NLLB model (Joulin et al., 2016), n=64,322 items.

The next step in text processing is therefore to translate all these non-English texts into English. To create 3D imagery from mesh models we employ a scripted Blender pipeline to create angle-varying render images, ingest in the Zenodo storage, and use it with the vision models.

Categorization
For categorization, the object categories were identified via the Europeana CHO Types Vocabulary for 3D content (Europeana, 2025) from the English or translated textual descriptions. A keyword-based search revealed n=5,798 results, with photographs and maps as top mentioned media, followed by buildings and sculpture as top content categories (Figure 4). No optimization to retrieve synonyms or word stems was applied at this training stage, nor were multiple nominations (e.g. photographs of buildings) eliminated.

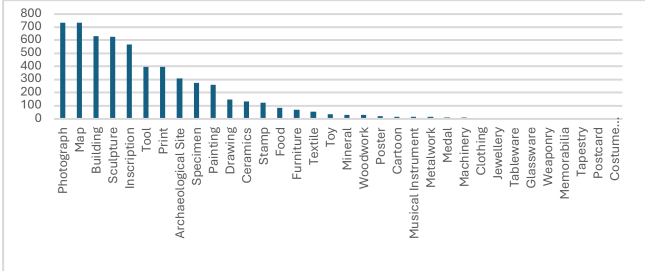

Figure 4. Identification of object categories via keywords in (translated) descriptions (n=5,798).

In an initial test we compared keywords assigned to images, descriptions, and mapped keywords. For the full Europeana dataset of 8,708 items we tested different approaches: Image classification was performed using CLIP-ViT-B/32 and Qwen3-VL-30B-A3B-Instruct. CLIP has been validated for as performing well for zeroshot classification (Li et al., 2023). Classification of descriptions and classification of not standardized concept labels both relied on MoritzLaurer/deberta-v3-large-zeroshot-v2.0. Classification of titles relied on the Qwen3-32B-MLX-4bit. All classifications were made using the Europeana classes.
To achieve greater precision in selecting architectural models, we trained a VGG-16 based classifier. The classifier was trained with 2,123 manually classified renderings from the Objaverse 3D dataset. For training, 1,699 files were used – including nine variants by data augmentation per file – and for validation we used 424 files. The trained classifier has 79.3% validation accuracy with 43.2% loss (Figure 5).

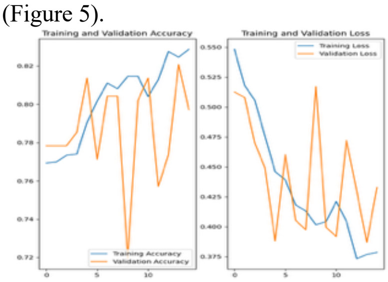

Figure 5. VGG-16-based training and validation accuracy and loss for classification 3D models of architectural exteriors/others.

The model was run for the Objaverse object dataset with 15,287 results labelled by the model with a confidence of more than 80%. From the classified objects 2,372 are classified as architectural exteriors with 12,915 classified as other content.

[10] https://huggingface.co/facebook/fasttext-language-identification
[11] https://huggingface.co/facebook/nllb-200-distilled-1.3B
[12] https://huggingface.co/sshleifer/distilbart-cnn-12-6
[13] https://huggingface.co/llava-hf/llava-v1.6-mistral-7b-hf
[14] https://huggingface.co/MoritzLaurer/deberta-v3-large-zeroshot-v2.0
[15] https://huggingface.co/Qwen/Qwen3-14B-MLX-4bit
[16] https://huggingface.co/Qwen/Qwen3-32B-MLX-4bit
[17] https://huggingface.co/sentence-transformers/clip-ViT-B-32
[18] https://huggingface.co/meta-llama/Meta-Llama-3-8B
[19] https://huggingface.co/deepseek-ai/DeepSeek-R1
[20] https://platform.deepseek.com
[21] https://spacy.io
[22] https://scikit-learn.org
[23] https://github.com/nicolas-dufour/plonk
[24] https://github.com/facebookresearch/OrienterNet?tab=readme-ov-file
[25] https://github.com/VicenteVivan/geo-clip

Geolocalization

With regards to geolocalization Figure 6 shows the results of benchmarked and tested LLMs, leveraging the Europeana dataset (Münster et al., 2025b). Promising results were achieved by using the DeepSeek-R1 671B parameter LLM with an 82% recognition rate for countries and a 46% rate for cities.

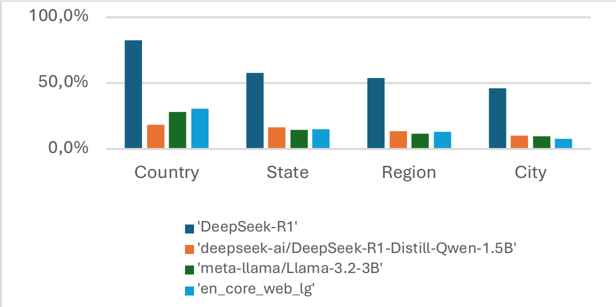

Figure 6. Matching text-based model retrieved to human-assigned location information exploiting the Europeana dataset (n=2,465).

An important finding was that the other models were significantly less accurate, with only 30% of countries correctly identified by the Spacy large English model (en_core_web_lg). To create additional manually validated data, the project partner PCSS has created a web-based browser add-on that allows the 3D model to be manually positioned on the terrain model. The resulting information on the position, scale, and orientation of the 3D model is stored in the record. The add-on integrates seamlessly with the Zenodo storage system. An alternative approach utilizes an LLM and VLM for rough geolocalization, and satellite imagery for fine geolocalization and pose estimation, as described in (Rigon et al., in press).

7. VIEWER INFRASTRUCTURES

Within the projects a set of viewers are under development (Figure 7). The project utilizes the DFG 3D Viewer as a flexible wrapper structure to accommodate various viewers (Münster et al., in press). Currently, models can be viewed in the DFG 3D Viewer with a choice of ten viewer integrations. The integrated viewers are ATON, 3D-HOP, UH4D Browser, Kompakkt, model-viewer, and the AIM Viewer for 3D models, as well as the PLY Point Cloud Viewer, XRWeb, and Holopyramid Viewer from the University of Jena for views of PLY models in a point cloud, Virtual Reality (VR) glasses and holopyramids (Figure 8). These viewers serve different usage scenarios such as scholarly annotation, XR exploration, and holopyramid visualizations, while maintaining a consistent link to the archived master asset.

Figure 7. DFG 3D Viewer integration examples of the Rooom Viewer (left) and the PCSS Viewer (right).

In the 3DBigDataSpace project numerous visualization stacks has been developed and integrated in the DFG 3D Viewer.

[26] https://artefact-iq.in-two.com/

The PCSS Viewer, developed as part of the 3DBigDataSpace project, is a standalone tool for viewing 3D objects. The PCSS Viewer was inspired by the 3D viewer implemented around 2019 for the development of dLibra software – a digital repository system (specifically, a custom implementation for the Digital Repository of the Scientific Institutes in Poland). The PCSS Viewer was built on a modern technology stack in a front-end architecture environment. It primarily combines the React and Three.js libraries. The tool currently includes basic functionalities for handling.glb/.gltf files, such as autorotation, object manipulation, displaying an object's mesh, changing colour, and displaying annotations. The PCSS Viewer has significant development potential, for example, the ability to expand its functional repertoire through iterative development. The 4D Browser and 4D City World viewers are used for virtual, augmented, and 2.5D visualization on mobile and desktop devices (Münster et al., 2024c). The 4D City application for mobile devices enables time-variant virtual 3D impressions of historic cities and the 4DBrowser enables browsing media collections on desktop systems.

The Extended Reality Visualizations include a VR viewer developed by the University of Jena to show 3D mesh models in the browser of VR glasses. Another application developed by the project partner Rooom AG makes it possible to create Extended Reality (XR) experiences to view specific 3D objects on mobile devices.

The Holopyramid Viewer is a browser-based software application capable of compiling the 4 synchronized corresponding views of the 3D model for any given model URL that would allow it to be visualized on a holopyramid device.

ArtefactIQ[26] is a web-based application that integrates the different viewers described above and enables the creation of multi-platform gamified experiences. Using a content management authoring interface and leveraging flexible templates, ArtefactIQ enables cultural heritage professionals and educators to reuse 3D models and datasets and build small games that are more engaging.

The UVigo application[27] was designed to enhance the cultural and tourist experience of pilgrimage routes through immersive technologies and AI.

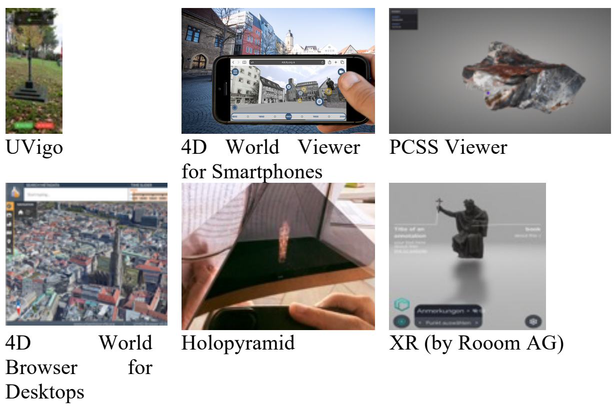

Figure 8. Application interfaces.

7. USE CASES

Museum exhibitions

The use of XR applications has a significant effect on museum learning (Xu et al., 2023), enjoyment (Dong et al., 2024) and engagement. To validate affective impact, we collaborated with the

[27] Available at https://camino-57345.web.app, https://play.google.com/store/apps/details?id=com.rurallure, October 2025.

Egypt Museum in Turin, the Alesia Museum, and MAC in Barcelona. The data pool was used to create a set of prototypes. Specific scenarios include (1) to validate affective and learning effects of remote and physical representations, (2) provide content in a remote environment vs on site, (3) test different ways of storytelling, e.g. related to multi-coding (Paivio, 2006), varying narratives, and customization (Shen et al., 2024). Several test settings were compiled to enable lab tests of different modes of interaction with content (Figure 9).

Image	3D View (navigable)	Holo-pyramid	VR

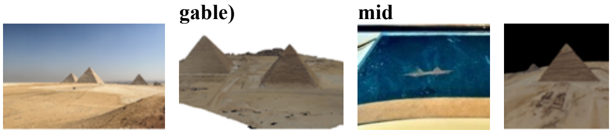

Figure 9. Pyramid experiences using the different viewer and datasets (Content: Plateau de Gizeh; generated by pierre391).

Education
Educational courses designed by students of art history and history teaching were offered to primary and secondary school students in Jena in the context of school and extracurricular working groups. These courses produced content for the 4D world viewers – e.g. 3D scans of city sculpture, textual descriptions of landmarks, and virtual city tours for children. To date 170 school students have participated in these educational courses (Münster et al., 2025a). The Hunt Museum in Limerick, Ireland, used the 3D tools to augment the existing learning material on *Life in the Bronze Age* for secondary students starting in November 2025. 3D models are used both individually, using the PCSS Viewer and its annotation capability, and within a gamified experience created with ArtefactIQ where students aim to recognize certain characteristics of museum artefacts (e.g. to date them). Initial sessions with students (58 respondents) revealed that using 3D models was a most distinctive and engaging experience and thus can be a strong intrinsic motivator. Further evidence that the 3D models were perceived as an important learning aid is provided by qualitative feedback, which stressed that the ability to view objects from all angles and to see details of internal structures was very useful in the students' leaning. Testing also revealed that the technical limitations of the devices used play an important role in the user experience (even if the pedagogical design is appreciated), and optimizations of the 3D models can reduce the negative impact. The positive initial results motivate further investigation of the approach and improvement of the underlying tools, e.g. to provide better content and explanations within the 3D viewer, to better use annyotations on the 3D models, and to improve the way in which feedback and progression is leveraged in the gamified experience.

Tourism
Virtual travel guides provide tourists with information about their route and surroundings. The UVigo application was designed to enhance the cultural and tourist experience of pilgrimage routes through immersive technologies and AI. The proposed architecture combines mobile development, real-time navigation, and AI-based content generation to create an interactive, multilingual environment that connects users with cultural routes (i.e., Camino Miñoto Ribeiro, Camino de Santiago, and Rome). In more detail, the application integrates AR, 3D visualization, and AI-driven enrichment tools for both textual and visual content. Other significant functionalities include real-time routing and emergency notifications that alert nearby users during their pilgrimage through the app's notification system. User tracking is also supported with real-time geolocation and alerts for points of interest, improving user safety and engagement.

8. RESULTS

Large-scale 3D datasets and versatile viewers offer various opportunities in education, tourism, and museums to illustrate and highlight information. They therefore enable the use of 3D in cultural heritage on a large scale.
To date, text-based approaches using captions or descriptions remain superior to image-based classification or geolocalization. Specifically, for geolocalization tasks, models with full-scale LLMs – exemplified with Deepseek-R1 – perform much better than smaller LLMs or transformer-based models.
Concerning provenance, most of the 3D cultural heritage content explored here stems from Mediterranean countries with buildings and sculptures named most frequently as object categories. This is in line with former investigations, in which the majority of digitized 3D objects were located in Italy, Spain and Greece (Münster, 2019).

9. CONCLUSIONS

The mentioned approaches are currently undergoing initial validation. The next step is to integrate promising candidate technologies into a modular processing pipeline for data ingestion in Zenodo and Europeana, which is currently in development. This pipeline includes additional components that are in development, such as a production-grade 3D remesher to adopt the level of detail of the 3D models, and integrating the different approaches as agents. After that, a test with the Objaverse XL Github and Thingiverse content of 8.9 M assets and a compiled dataset including 1.2 M segmented LiDAR & OSM tiles is planned.
Finally, the use cases are currently helping to validate the datasets and viewer technologies in real-world scenarios.

ACKNOWLEDGEMENTS

The research which this paper is based on was carried out in the EU projects 3DBigDataSpace (Grant No. 101173385) and MetaMuseum (Grant No. 101132488). We thank the project partners in both projects for their input into and feedback on the investigations and the provision of use cases described in the article.

REFERENCES

Aiger, D., Allen, B., and Golovinskiy, A., 2017. Large-Scale 3D Scene Classification With Multi-View Volumetric CNN, ArXiv, 1712.09216, https://doi.org/10.48550/arXiv.1712.09216.

Bajena, I., Dworak, D., Kuroczyński, P., Smolarski, R., and Münster, S., 2022. DFG-3D-Viewer – Development of an infrastructure for digital 3D reconstructions, Proceedings of the DH2022 Conference Abstracts, DH2022 Local Organizing Committee, Tokyo, Japan, 25–29 July 2022.

Blettery, E., Fernandes, N., and Gouet-Brunet, V., 2021. How to Spatialize Geographical Iconographic Heritage, Proceedings of the 3rd Workshop on Structuring and Understanding of Multimedia heritAge Contents, Virtual Event, China, 10.1145/3475720.3484444.

Cepeda, V. V., Nayak, G. K., and Shah, M., 2023. GeoCLIP: clip-inspired alignment between locations and images for effective worldwide geo-localization, Proceedings of the 37th International Conference on Neural Information Processing Systems, New Orleans, LA, USA.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-2-W12-2026-319-2026

Chang, A. X., Funkhouser, T., Guibas, L., Hanrahan, P., Huang, Q., Li, Z., Savarese, S., Savva, M., Song, S., and Su, H., 2015. Shapenet: An information-rich 3d model repository, arXiv, 1512.03012.

Deitke, M., Liu, R., Wallingford, M., Ngo, H., Michel, O., Kusupati, A., Fan, A., Laforte, C., Voleti, V., Gadre, S. Y., VanderBilt, E., Kembhavi, A., Vondrick, C., Gkioxari, G., Ehsani, K., Schmidt, L., and Farhadi, A., 2023. Objaverse-XL: A Universe of 10M+ 3D Objects, arXiv, 2307.05663, https://doi.org/10.48550/arXiv.2307.05663.

Dong, H., Cai, Y., and Wei, Q., 2024. The Impact of VR Museum Exhibition Design on User Emotional Experience: A Meta-Analysis, Proceedings of the 2024 International Conference on Cloud Computing and Big Data, Dali, China, 10.1145/3695080.3695105.

Dufour, N., Picard, D., Kalogeiton, V., and Landrieu, L., 2024. Around the World in 80 Timesteps: A Generative Approach to Global Visual Geolocation, arXiv, 1309.2434.

European Commission, 2022. Study on quality in 3D digitisation of tangible cultural heritage: mapping parameters, formats, standards, benchmarks, methodologies, and guidelines. VIGIE 2020/654 Final Study Report.

Europeana, 2025. Europeana CHO Types Vocabulary v2.

Fiorucci, M., Khoroshiltseva, M., Pontil, M., Traviglia, A., Del Bue, A., and James, S., 2020. Machine Learning for Cultural Heritage: A Survey, Pattern Recognition Letters, 133, 102-108, 10.1016/j.patrec.2020.02.017.

Sketchfab, 2022. Over 100,000 Cultural Heritage Models on Sketchfab: https://sketchfab.com/nebulousflynn/collections/over-100000-cultural-heritage-models-on-sketchfab, accessed 29 January 2022.

Fung, N., Schoueri, K., and Scheibler, C., 2021. Pure 3D: Comparison of Features Available on Aton, Smithsonian Voyager, 3DHOP, Kompakkt and Virtual Interiors (Technical Report), Maastricht University, Maastricht, The Netherlands.

Gominski, D., Poreba, M., Gouet-Brunet, V., and Chen, L., 2019. Challenging Deep Image Descriptors for Retrieval in Heterogeneous Iconographic Collections, Proceedings of the 1st Workshop on Structuring and Understanding of Multimedia heritAge Contents, Nice, France, 10.1145/3347317.3357246.

Haas, L., Skreta, M., Alberti, S., and Finn, C., 2024. PIGEON: Predicting Image Geolocations, arXiv, 2307.05845v6.

Jaillot, V., 2020. 3D, temporal and documented cities: formalization, visualization and navigation.

Joulin, A., Grave, E., Bojanowski, P., and Mikolov, T., 2016. Bag of Tricks for Efficient Text Classification, arXiv, 1607.01759.

Kulkarni, P. P., Nayak, G. K., and Shah, M. , 2024. CityGuessr: City-Level Video Geo-Localization on a Global Scale, Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part LXIII, Milan, Italy, 10.1007/978-3-031-73036-8_17.

Li, X., Wen, C., Hu, Y., and Zhou, N., 2023. RS-CLIP: Zero shot remote sensing scene classification via contrastive vision-language supervision, International Journal of Applied Earth Observation and Geoinformation, 124, 103497, https://doi.org/10.1016/j.jag.2023.103497.

Lin, C., Liu, H., Lin, Q., Bright, Z., Tang, S., He, Y., Liu, M., Zhu, L., and Le, C., 2025. Objaverse++: Curated 3D Object Dataset with Quality Annotations, arXiv, 2504.07334.

Maiwald, F. , 2022. A window to the past through modern urban environments — Developing a photogrammetric workflow for the orientation parameter estimation of historical images, 10.13140/RG.2.2.19627.52004.

Maiwald, F., Lehmann, C., and Lazariv, T., 2021. Fully Automated Pose Estimation of Historical Images in the Context of 4D Geographic Information Systems Utilizing Machine Learning Methods, ISPRS International Journal of Geo-Information, 10, 748.

Morelli, L., Bellavia, F., Menna, F., and Remondino, F., 2022. Photogrammetry now and then - from hand-crafted to deep-learning tie points, ISPRS - International Archives of the Photogrammetry Remote Sensing and Spatial Information Sciences, XLVIII-2/W1-2022, 163-170, 10.5194/isprs-archives-XLVIII-2-W1-2022-163-2022.

Münster, D. L., Münster, S., and Dietz, R., 2025a. Heritage education projects by university students for pupils: Digital and data-driven learning in a humanities teaching-learning lab, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-M-9-2025, 1037-1042, 10.5194/isprs-archives-XLVIII-M-9-2025-1037-2025.

Münster, S., 2023. Advancements in 3D Heritage Data Aggregation and Enrichment in Europe: Implications for Designing the Jena Experimental Repository for the DFG 3D Viewer, Applied Sciences, 13, 9781.

Münster, S., Lehmann, C., Lazariv, T., Maiwald, F., and Karsten, S. , 2021., Toward an Automated Pipeline for a Browser-Based, City-Scale Mobile 4D VR Application Based on Historical Images, Research and Education in Urban History in the Age of Digital Libraries, Cham, 106-128.

Münster, S., Bruschke, J., Dworak, D., Komorowicz, D., Rajan, V., and Ukolov, D., 2024a. 4D geo modelling from different sources at large scale, ACM Multimedia Melbourne.

Münster, S., Bruschke, J., Rajan, V., Komorowicz, D., Preßler, R., and Ukolov, D., 2025b. 4D World Viewers as Multi-user Content Management Systems, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci., XLVIII-M-9-2025, 1043-1050, 10.5194/isprs-archives-XLVIII-M-9-2025-1043-2025.

Münster, S., Kuroczynski, P., Mosch, M., Bajena, I., Beck, C., Weigelt, M., and Dworak, D., in press. Implementing a distributed infrastructural workflow for 3D Models in Germany. Achievements of the DFG 3D-Viewer Project's Second Funding Phase, in: Research and Education in Urban History in the Age of Digital Libraries 2025, edited by: Münster, S., Kuroczyński, P., and Apollonio, F., Springer, Cham.

Münster, S., Maiwald, F., Lenardo, I. d., Henriksson, J., Isaac, A., Graf, M., Beck, C., and Oomen, J., 2024b. Artificial Intelligence for Digital Heritage Innovation. Setting up a R&D roadmap for Europe, MDPI Heritage, 7, 794–816.

Münster, S., Maiwald, F., Bruschke, J., Kröber, C., Sun, Y., Dworak, D., Komorowicz, D., Munir, I., Beck, C., and Münster, D. L., 2024c. A Digital 4D Information System on the World Scale: Research Challenges, Approaches, and Preliminary Results, Applied Sciences, 14 ,1992.

Orzechowski, M., Opioła, Ł., Martínez, I. L., Ioannides, M., Panayiotou, P. N., Dutka, Ł., Słota, R. G., and Kitowski, J., 2025. Integrated data, metadata, and paradata management system for 3D Digital Cultural Heritage objects: Workflow automation, federated authentication, and publication, Future Generation Computer Systems, 107964, https://doi.org/10.1016/j.future.2025.107964.

Paivio, A., 2006. Dual Coding Theory and Education (Draft), Pathways to Literacy Achievement for High Poverty Children," The University of Michigan School of Education, 29. 9.- 1. 10. 2006.

Pramanick, S., Nowara, E. M., Gleason, J., Castillo, C. D., and Chellappa, R. 2022. Where in the World Is This Image? Transformer-Based Geo-localization in the Wild, Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXVIII, Tel Aviv, Israel, 10.1007/978-3-031-19839-7_12.

Radovic, M., Adarkwa, O., and Wang, Q. S., 2017. Object Recognition in Aerial Images Using Convolutional Neural Networks, Journal of Imaging, 3, 10.3390/jimaging3020021.

Rigon, S., Farella, E. M., Morelli, L., Bertolasi, G., Remondino, F., and Münster, S., in press. 3DGeoRef: an automated framework for georeferencing heritage 3D models, Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.

Sarlin, P.-E., Cadena, C., Siegwart, R., and Dymczyk, M., 2019. From Coarse to Fine: Robust Hierarchical Localization at Large Scale, ArXiv, 1812.03506.

Sarlin, P.-E., DeTone, D., Yang, T.-Y., Avetisyan, A., Straub, J., Malisiewicz, T., Bulo, S. R., Newcombe, R., Kontschieder, P., and Balntas, V., 2023. OrienterNet: Visual Localization in 2D Public Maps with Neural Matching, CVPRW.

Sattler, T., Maddern, W., Toft, C., Torii, A., Hammarstrand, L., Stenborg, E., Safari, D., Okutomi, M., Pollefeys, M., Sivic, J., Kahl, F., and Pajdla, T., 2018. Benchmarking 6DOF Outdoor Visual Localization in Changing Conditions, 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 18-23 June 2018, 8601-8610, 10.1109/CVPR.2018.00897.

Schaffland, A., Bui, T. H., Vornberger, O., and Heidemann, G. 2020. New Interactive Methods for Image Registration with Applications in Repeat Photography, Proceedings of the 2nd Workshop on Structuring and Understanding of Multimedia heritAge Contents, Seattle, WA, USA, 10.1145/3423323.3425749.

Schindler, G. and Dellaert, F., 2012. 4D cities: analyzing, visualizing, and interacting with historical urban photo collections, Journal of Multimedia, 7, 124-131.

Shen, J., Mire, J., Park, H. W., Breazeal, C., and Sap, M., 2024. HEART-felt Narratives: Tracing Empathy and Narrative Style in Personal Stories with LLMs.

Singh, A. and Aneja, S., 2024. NewsCaption: Named-Entity aware Captioning for Out-of-Context Media, arXiv, 2403.12618.

Storeide, M. S. B., George, S., Sole, A., and Hardeberg, J. Y., 2023. Standardization of digitized heritage: a review of implementations of 3D in cultural heritage, Heritage Science, 11, 249, 10.1186/s40494-023-01079-z.

Tung, J., Chou, G., Cai, R., Yang, G., Zhang, K., Wetzstein, G., Hariharan, B., and Snavely, N., 2024. MegaScenes: Scene-Level View Synthesis at Scale, https://arxiv.org/abs/2406.11819, 2406.11819.

Vivanco, V., Nayak, G. K., and Shah, M., 2023. GeoCLIP: Clip-Inspired Alignment between Locations and Images for Effective Worldwide Geo-localization, Advances in Neural Information Processing Systems.

Wang, Z., Xu, D., Khan, R., Lin, Y., Fan, Z., and Zhu, X., 2024. LLMGeo: Benchmarking Large Language Models on Image Geolocation In-the-wild, ArXiv, 2405.20363, 10.48550/arXiv.2405.20363.

Wu, Y., Shi, L., Liu, H., Liao, H., Qiu, L., Yuan, W., Gu, X., Dong, Z., Cui, S., and Han, X., 2024. MVImgNet2.0: A Larger-scale Dataset of Multi-view Images, ACM Trans. Graph., 43, Article 173, 10.1145/3687973.

Xu, S., Zhang, C., Fan, L., Meng, G., Xiang, S., and Ye, J., 2024. AddressCLIP: Empowering Vision-Language Models for City-Wide Image Address Localization, Computer Vision – ECCV 2024: 18th European Conference, Milan, Italy, September 29–October 4, 2024, Proceedings, Part XXVIII, Milan, Italy, 10.1007/978-3-031-73390-1_5.

Xu, W., Dai, T.-T., Shen, Z.-Y., and Yao, Y.-J., 2023. Effects of technology application on museum learning: A meta-analysis of 42 studies published between 2011 and 2021, Interactive learning environments, 31, 4589-4604.