Cultural Heritage 3D Object Management with Integrated Automation Workflows

Michał Orzechowski[1,2], Łukasz Opioła[1], Ignacio Lamata Martínez[3], Marinos Ioannides[4], Panayiotis N. Panayiotou[4], Renata G. Słota[2], Łukasz Dutka[1], and Jacek Kitowski[1,2]

[1] Academic Computer Centre Cyfronet AGH, Krakow, Poland
Email: {m.orzechowski,l.opiola,lukasz.dutka}@cyfronet.pl
[2] AGH University of Krakow, Faculty of Computer Science, Poland
Email: {rena,kito}@agh.edu.pl
[3] EGI Foundation, Amsterdam, The Netherlands
Email: ignacio.lamata@egi.eu
[4] Digital Heritage Lab, Cyprus University of Technology, Cyprus
Email: {p.panayiotou,marinos.ioannides}@cut.ac.cy

Abstract. The complexity of high-quality 3D digitised cultural heritage objects creates challenges for existing data management systems as they need to develop metadata management and processing capabilities to provide semantic insight into the interconnectivity of data that constitute cultural heritage objects. To address these challenges, we propose a data and metadata management system, together with the federated authentication and authorisation mechanism, and an integrated system for designing and executing automated workflows that facilitate the processing of both data and metadata. The solution is evaluated with a 3D digitised cultural object of *Lambousa Fishing Boat* and presents the complete process from data upload to publication.

Keywords: data management · workflow processing · automation workflows · metadata management · paradata · data repository · cultural heritage · cultural heritage objects · 3D digitalisation

1 Introduction

Over past years, the European Union has allocated significant resources [11][8][6] towards the digitalisation of Cultural Heritage assets (CH). This funding has been instrumental in supporting CH entities like museums, archives, and libraries in digitising, archiving, and preserving their collections for research, innovation, and educational purposes. This investment period has been pivotal for research, innovation, education, and cross-sectoral development in the European CH sector. In parallel, the last years have witnessed significant private, governmental, and non-governmental initiatives in 2D and 3D digitalisation from entities such as Google [14], CyArk [7] or the Smithsonian Institute [23].

The advancements in ICT have catalysed an exponential increase in digital content production, with the digital universe expected to expand tenfold in the coming years. Significant strides in 3D digitalisation have enhanced the accessibility of European CH for multiple purposes. High-quality 3D scans aid in archaeological and engineering conservation efforts, while medium-quality data find applications in the creative industry.

This paper presents our solution, which enables 3D Cultural Heritage Objects (CHO) management, processing, and integration with external cultural heritage digital collections. The proposed implementation combines data access and management functionality, metadata management, processing capabilities of the integrated workflow automation platform, and the federated authentication and authorisation mechanism. The novelty of this paper is the introduction of automated workflows for the management and processing of CHO's data, metadata, and paradata.

The paper is organised as follows. Section 2 discusses the problem statement and related work in the area of digital infrastructures of management and processing of 3D objects. Section 3 describes the integrated automation workflow implementation as part of the Onedata platform. Section 4 presents our solution for CHO processing, in particular the federated authentication and authorisation mechanism, data metadata processing, and CHO publication. Section 5 provides an evaluation of the solution, describing the comprehensive processing of a complex 3D cultural heritage object. Finally, Section 6 concludes the paper.

2 Problem Statement and Related Work

The 3D digitalisation of movable and immovable CHO is an inherently complex, multi-stage process. Different parameters such as infrastructure hardware, software, personnel, etc., could influence this process and yield different results. Furthermore, there are unique documentation challenges for the CHOs within each organisational category, and the capabilities of the recording hardware, associated processing software, production methods and visualisation systems are continuously evolving. With the rapid expansion of digitalisation projects and advancements in laser scanning systems, the integration of photogrammetric imagery, high-resolution still images, renderings, and animations has become more prevalent.

Consequently, the need for CHO repositories has become increasingly clear. CH institutions can benefit significantly from such repositories, as they allow secure sharing of the scan data with administrators, clients, scholars, experts, and contractors worldwide. This setup enables CH sites to lease computing resources, software, and systems on an as-needed basis via the Internet, thus reducing the expenses associated with traditional IT infrastructure. Additionally, there is a growing need for a cloud-based platform dedicated to the semantic enrichment and visual analysis of 3D models within the CH sector. Cloud infrastructures must evolve to become all-encompassing solutions that offer the necessary ICT support to host high-quality content and user-friendly graphical interfaces.

From an infrastructure perspective, the primary artefacts of the 3D digitalisation process are the 3D model itself and the metadata and paradata [2] associated with it. Together, they constitute a CHO. Comprehensive and integrated management of all these components poses challenges for data management systems:

- archival and retrieval of resources that link various parts of a 3D object to its main structure,
- tools for visual data analysis, data annotation, modification, validation, and importation/exportation of paradata and metadata,
- advanced functionalities for importing and exporting 3D models in all available 2D and 3D formats, addressing related intellectual property rights (IPR) issues,
- the capability to embed additional information into the 3D model structure, such as multimedia-linked data (audio, video, text, image),
- single sign-on, federated authentication and authorisation mechanism to facilitate integration with research and cultural institutions,
- distributed data management to allow federated management and contribution of infrastructure resources,
- metadata-aware data management,
- integrated, declarative data processing capabilities for creation of 3D models processing pipelines as well as metadata validation and processing,
- sharing and publication of CHOs (particular models accompanied by metadata) as Open Data, or publishing to CH aggregators such as Europeana.

Such requirements call for designing a system that can handle the complexity of CHO objects and allow management and processing of both the data and the metadata, encompassing the whole process from the data ingestion to CHO publication.

The 3D digitalisation of CHO is an inherently complex multi-stage process [15], and the final result consists of 3D models, metadata, and associated paradata. Moreover, different 3D digitised objects might require dedicated processing methods and procedures. This requires designing dedicated workflows for processing each type (or even each instance) of the 3D digitised object.

A large number of infrastructures currently focus on 3D models, with recent reports [5][4] summarising existing platforms, repositories [1][20][19], visualisation frameworks and formats [13]. Digital infrastructures, services, and tools represent distinct yet interdependent components within the digital ecosystem. Digital tools are specialised software applications engineered for specific tasks, such as the generation of 3D models or the transformation of metadata. Conversely, a digital service constitutes a managed provision of a tool to external users, while digital infrastructures encompass one or more services and/or facilities that are required for the operation and delivery of digital capabilities.

In [18], authors distinguish four types of infrastructures for CHO and 3D models: data repositories, data aggregators, 3D data viewers, and virtual research environments. Repositories are collections of 3D models that are stored and made available for secondary use, the most widely used 3D repository globally being Sketchfab [16]. Data aggregators do not store 3D models directly, but instead compile databases to prepare combined data sets for data processing. For example, Europeana [12] acts as a data aggregator by compiling its collection from data in national and regional libraries. 3D data viewers employ computer graphics to dynamically visualise and interact with 3D models, enabling functionalities such as rotation and zooming, e.g., Kompakkt [17]. Virtual research environments are web-based information systems that provide a comprehensive working environment for researchers, including various tools for analysis and comparison [3].

Onedata is a global data management system, providing easy access to distributed storage resources and supporting use cases from personal data management to data-intensive scientific computations. It meets the most crucial of the above requirements related to the management of distributed data and metadata, data publication, integration with multiple identity providers, storage resource management and integrated data processing. The system has already been used to provide distributed access and data processing of more than 4PB of archival cultural data sets from Polish national museums [22]. It has also been used as a data aggregation and processing hub for data originating from different life-science experimental methods [24].

Hence, in our solution, Onedata was chosen as a data and metadata management platform, together with the EGI Check-in[9] system for providing federated authorisation for handling complex data and metadata of CHO via the introduction of automation workflows as an integrated part of the data management system.

3 Automation workflows with Onedata

Integrated automation workflows (AW) within the Onedata system provide a flexible and efficient means to automate tasks, allowing users to compose sequences of actions and execute them on selected data collections. While Onedata orchestrates workflow execution, users are responsible for implementing the logic of individual data processing functions, known as lambdas, which dictate how data is manipulated, transformed, or analysed. This flexibility empowers users to tailor their processing steps to suit their specific requirements without imposing restrictions on the nature of the logic.

The term *integrated* signifies that these workflows are seamlessly run by Onedata's built-in automation engine. This integration enables lambdas to directly access distributed user data and process or modify it, with data distribution abstracted by the Onedata layer. Regardless of the data's location, whether local or remote, it is delivered on-demand to the lambdas. These lambdas have access to a wide range of Onedata APIs, including Quality of Service rules, custom metadata management, public sharing, and archiving, ensuring comprehensive data management capabilities. Additionally, workflows are executed in a containerized fashion on Kubernetes, facilitating on-demand and scalable execution of tasks. The execution layer is abstracted out — while the current implementation uses OpenFaaS [21] to run the lambda containers, other task execution platforms/frameworks can be easily plugged in.

The main concepts related to automation workflows in Onedata are: **inventories**, **lambdas**, and **workflow schemas**. An inventory is a registry of lambdas and workflow schemas that can be made accessible to multiple users and user groups, giving them a way to share and reuse common definitions. A lambda is essentially a task with arbitrary logic that takes input arguments and runs a user-implemented handler to produce results. The specification of arguments and results are included in the lambda definition, together with their data types. During execution, input and output data is encoded to JSON during transmission to and from the handler, making it easy to implement in any programming language.

Workflow schemas can be perceived as blueprints for data processing pipelines — cf. Fig. 1. They run a series of lambdas on a set of inputs to produce outputs and, possibly, side effects (e.g. newly created files). Intermediate products of processing steps can be used as inputs for consecutive tasks — to that end, the so-called **stores** are used. They hold collections of typed data and act as sinks for the results produced by lambdas. Stores can also be used to hold the user input required for starting an execution or the final results of the whole run. A workflow schema consists of one or more **lanes** that are processed sequentially. Each lane has one or more lambdas and one designated store it iterates over, running all the lambdas for each element in the store. Within a lane, the lambdas are executed sequentially, unless they are placed in a **parallel box**, allowing them to be run in parallel.

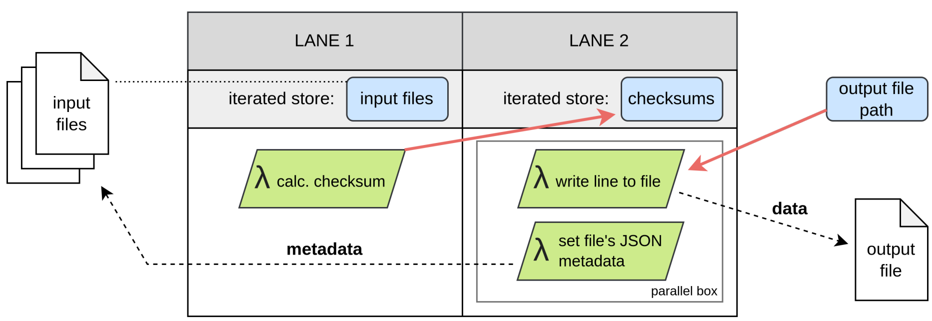

Fig. 1. Workflow schema example, showing its basic anatomy.

The example in Fig. 1 shows a workflow schema that takes a set of input files and calculates their checksums, storing them as JSON metadata of corresponding input files and writing a summary to an output file. The stores *input files* and *output file path* hold the initial user input required to run the workflow. The *checksums* store serves both as an intermediary container for data and a final artefact of the execution, storing `{file, checksum}` pairs. In the second lane, this store is iterated over and for every pair, a new line to the summary output file is written and a JSON metadata entry with the checksum is assigned to the input file.

4 Cultural Heritage Objects Processing

The 3D digitalisation of CHO is an inherently complex multi-stage process, and the final result consists of 3D models, metadata, and associated paradata. Moreover, different 3D digitised objects might require dedicated processing methods and procedures. This requires designing dedicated workflows for processing each type (or even each instance) of the 3D digitised object.

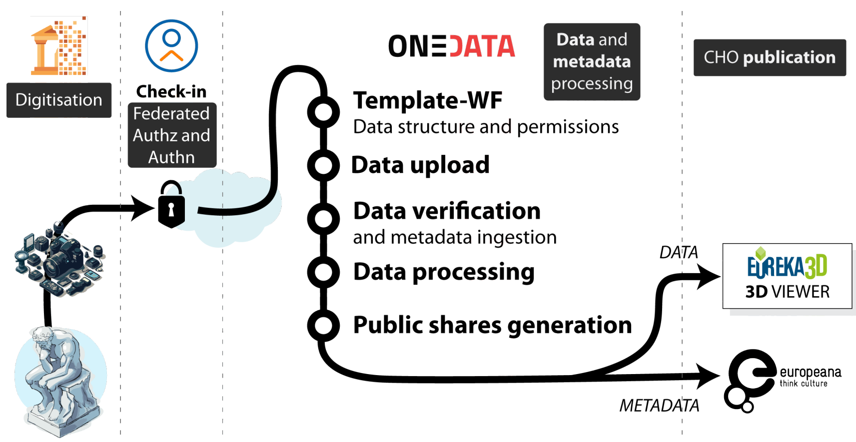

Fig. 2. Generic workflow processing of CHO processing.

Although the responsibility for the process of performing 3D digitalisation rests solely on the entity of the content provider with access to the object and digitalisation equipment, the following digital processing can be standardised and documented within the data processing platform. This allows for the reproducibility of the 3D object processing and enables reusing common parts of the process for different 3D objects. Fig. 2 shows the proposed generic workflow for a CHO processing starting with the digitalisation of the 3D model, followed by the authenticated upload of the digitalisation results to the Onedata platform, processing of the results using dedicated automated workflows, and finally creating a public CHO record within Onedata, which can be accessed in 3D Viewer and embedded in Europeana.

The following subsections describe each part of the generic workflow, excluding the digitalisation, in which results are treated as input for the workflow.

4.1 Federated Authentication and Authorisation

The underlying set of cloud resources used by the various CH institutions, including servers, storage, and services, should be protected against unauthorised access. This is critical to ensure the trust and integrity of the valuable digital assets managed in Cultural Heritage (mainly data, metadata, and paradata), and it is primarily achieved through the use of two mechanisms: authentication and authorisation. Those two mechanisms are managed by EGI Check-in to securely upload the results of the digitalisation process.

The EGI Check-in service provides identity and access management components that facilitate users to access community services and resources. EGI Check-in acts as an intermediate system connecting users, authentication servers, and services, offering users authenticated access to services and enabling single sign-on. It supports widely adopted standards and open technologies, including OIDC, OAuth, SAML, and X.509, which facilitates interoperability and integration with existing AAI services (responsible for managing the authentication and authorisation) of other Research Infrastructures and Research Communities. EGI Check-in has the following features:

- provides increased productivity, with mechanisms such as single sign-on, and security,
- accepts multiple federated authentication sources using different technologies,
- is federated in eduGAIN[5] as a service provider, publishing REFEDS RnS[6] and Sirtfi[7] compliance,
- provides a graphical user interface for user registration and management, which allows identity unification through account linking, as well as an API for programmatic user management,
- can combine user attributes originating from various authoritative sources and deliver them to the connected service providers transparently. This process is conducted according to the GDPR,
- is a member of the EOSC[8] Access Federation.

In practical terms, EGI Check-in makes it possible to identify users who wish to access applications or data and to protect these resources from unauthorised access. This is done transparently to the user in a federated environment, which interconnects different cloud providers offering different resources in distinct geographic locations.

4.2 Data and Metadata Processing

The processing starts with a member of the content provider group from one of the CH institutions uploading the artefacts of the 3D digitalisation process:

[5] https://edugain.org
[6] https://refeds.org
[7] https://refeds.org/sirtfi, https://aarc-project.eu/policies/sirtfi
[8] https://eosc.eu

the 3D model and the associated metadata. However, before that happens, the content provider executes a dedicated CHO initialisation AW (*Template-WF* workflow) creating a directory structure within the dedicated CH repository. The appropriate access policies are applied to the directory structure, allowing access only to the content provider. Now, the content provider can upload the files containing the 3D model and the metadata to designated directories.

Then, the AW for data and metadata *validation* is executed. It validates the content of the 3D models and the files containing metadata. This step can be tailored to any 3D file format and metadata scheme, ensuring that uploaded data is correct. The metadata is then ingested into the Onedata platform, and the 3D files are annotated with it. This marks the logical creation of the CHO within the Onedata platform, allowing the CHO and its files to be searched in Onedata using the keywords in metadata, and later use this metadata to expedite the publication of the CHO. The 3D models that constitute the CHO are often large; this requires one to generate less complex (lower resolution textures, lower number of vertices) models of smaller size, suitable for later publication. This marks the creation of several derivative CHO objects, each containing the processed 3D model and the same metadata and paradata as the original CHO.

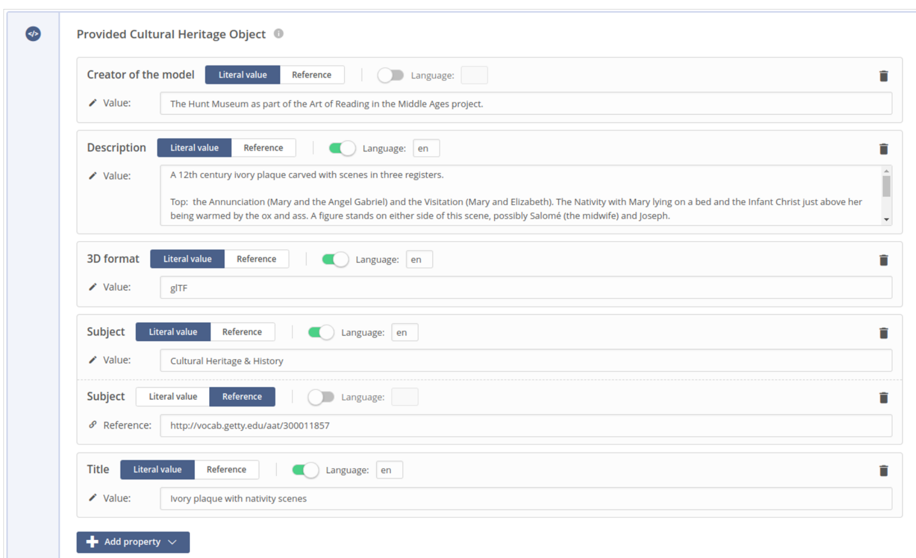

Fig. 3. Publishing CHO with Europeana Data Model.

In the last step, after verifying that the CHO object and its derivatives are correct, they can be published as single or separate public unlisted shares (only the person with the dedicated link can access it). Each share is annotated with the CHO metadata, and the additional paradata of the CHO processing workflow is added. This share can be ingested by the 3D Viewer, which extracts the 3D model or the list of its derivatives, allowing the end user to view it in the web browser.

4.3 Cultural Heritage Objects Publication

Onedata supports publishing shares as Open Data collections that are assigned open-access identifiers such as DOI or PID. Such Open Data shares can be annotated with Dublin Core metadata and advertised using OAI-PMH protocol by Open Data providers. However, CH communities in Europe strongly prefer to publish the CHO in the Europeana service.

That required implementation of support for European Data Model[9] (EDM) metadata schema for publishing Onedata shares. As part of Onedata, the dedicated graphical interface (see Fig. 3) for the EDM was implemented to assist with the publication of CHO. The required fields are auto-filled using the CHO metadata ingested into Onedata, to be manually approved by the content provider. In addition, a link to the 3D Viewer is provided to enable Europeana to embed the CHO object together with its interactive representation.

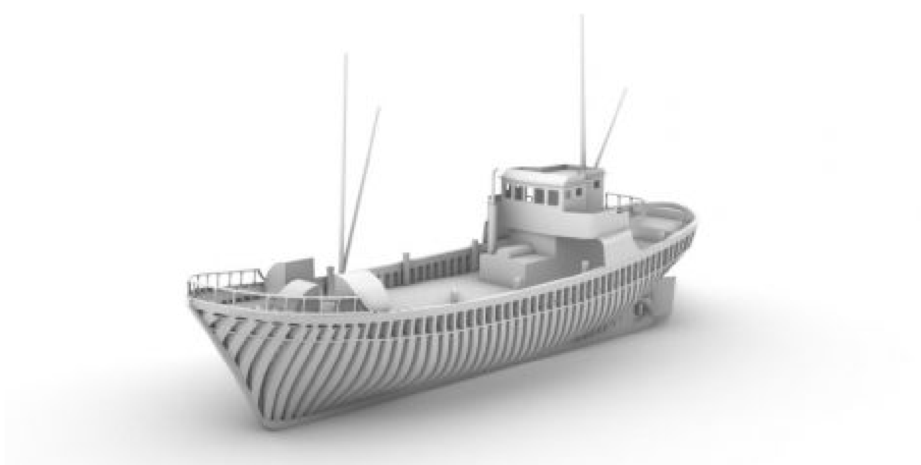

Fig. 4. 3D model of the *Lambousa Fishing Boat.*

5 Evaluation

The described automated workflows for the CH domain were developed as part of the EUreka3D project [10]. For their evaluation, we used the *Lambousa Fishing*

[9] https://pro.europeana.eu/page/edm-documentation

Boat[10], which was also digitised as part of EUreka3D by the Cyprus University of Technology.

The digitisation process yielded the 3D model (see Fig. 4) of the *Lambousa Fishing Boat* and accompanying files (of multiple data formats, eg.: `e57`, `obj`, `jpg`, `stl`, `3dm`) and was enriched with metadata detailed information about the boat itself and vast amount of paradata describing software and hardware that was used for the digitalisation, the weather conditions and much more.

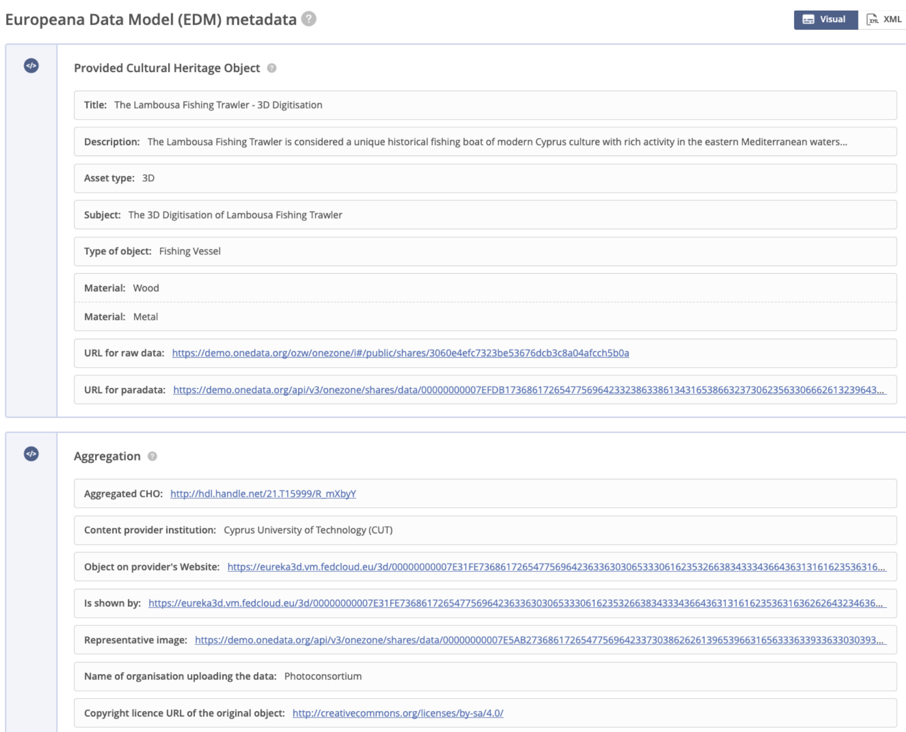

Fig. 5. Europeana data model with a subset of *Lambousa Fishing Boat* metadata.

Thanks to the Cyprus University of Technology being part of the eduGAIN network, the content provider could easily authenticate using EGI Check-in and log in to the Onedata platform. The content provider executed the templating AW prepared the directory structure and uploaded the digitisation results.

The data verification process especially designed to handle file types resulting from the boat digitalisation validated their correctness and extracted their metadata ingesting it into Onedata and attaching it to the files. The metadata and paradata of the 3D model were validated against the EDM scheme and the

[10] https://erachair-dch.eu/cutting-edge-digital-preservation-of-the-historic-lambousa-fishing-vessel

3D model files were annotated with it. The result was the *Lambousa Fishing Boat* CHO that can be managed by the Onedata platform.

The dedicated 3D processing AW converted the original `obj` and `mtl` files into multiple `glb` files (compatible with web-based 3D Viewer) derivatives of varying complexity and size, resulting in derivatives of the original CHO. Next, the public unlisted shares were created for each CHO allowing access to the 3D models from the 3D Viewer.

Finally, the CHO of the EDM was created (using the ingested CHO metadata from the Onedata system) and manually validated by the content provider, with the result depicted in Fig. 5. The published model was then registered in the Europeana database[11]. As a result, the 3D Viewer with the live 3D model of the *Lambousa Fishing Boat*, the metadata and the paradata were embedded in Europeana, using EDM scheme, and became available to the public.

6 Conclusions

The linking of historical sources of different genres, their digitised data, digital research artefacts, results, and associated metadata, paradata has been the focus of numerous projects. The recent rapid development of technologies allowing for the detailed 3D digitalisation of CHO presented new challenges in the standardisation and complexity management of CHO. This creates challenges for existing data management systems as they need to improve metadata management capabilities to develop processes for leveraging metadata to provide better data management and semantic insight into data interconnectivity — making the metadata the first-class citizen in the data management system.

Because of that, we propose a flexible system of automated workflows in which logical architecture is tightly integrated with the data management system, and at the same time its execution layer can leverage any type of job processing backend. Such a system can be used to design dedicated workflows for handling each step of CHO processing. What can foster the creation of workflow libraries that can be reused when composing dedicated workflows for different types of CHO. Moreover, such workflows are implemented in a declarative manner and combined with archiving and versioning capabilities of the data management system, allow the reproducibility of the whole CHO processing workflow. The versioning capability of the data management system could serve as a basis for designing a system for the management of 4D CHO with an integrated 4D Viewer, allowing to witness the evolution of 3D CHO over time.

Acknowledgements

This work is co-financed in part supported by the program of the Ministry of Science and Higher Education entitled *PMW* and under the program entitled *International Co-financed Projects* (projects no. 5398/DIGITAL/2023/2 and 5399/DIGITAL/2023/2).

[11] https://pro.europeana.eu/post/discover-how-the-eureka3d-project-supports-3d-in-the-data-space-for-cultural-heritage

JK and RGS are grateful for support from the subvention of the Polish Ministry of Education and Science assigned to AGH University of Krakow.

References

1. Benardou, A., Champion, E., Dallas, C., Hughes, L.M.: Cultural heritage infrastructures in digital humanities. Routledge London (2018)
2. Bentkowska-Kafel, A., Denard, H., Baker, D.: Paradata and Transparency in Virtual Heritage. Ashgate Publishing Company, USA (2012)
3. Candela, L., Castelli, D., Pagano, P.: Virtual research environments: an overview and a research agenda. Data Science Journal pp. GRDI75–GRDI81 (2013). https://doi.org/10.2481/dsj.GRDI-013
4. Champion, E., Rahaman, H.: Survey of 3d digital heritage repositories and platforms. Virtual Archaeology Review **11**(23), 1–15 (2020). https://doi.org/10.4995/var.2020.13226
5. Cieslik, E.: 3D digitization in cultural heritage institutions guidebook. Univeristy of Maryland: Baltimore, MD, USA (2020)
6. Culture in the EU National Recovery and Resilience Plan, https://cultureactioneurope.org/wp-content/uploads/2021/11/NRRPs_analysed_digital.pdf, accessed on 10/05/2024
7. CyArk - 3D recording technologies for the preservation and cultural heritage, https://www.cyark.org, accessed on 10/04/2024
8. Digital Europe Programme: Commission opens calls to invest in data spaces for tourism and cultural heritage, https://digital-strategy.ec.europa.eu/en/news/digital-europe-programme-commission-opens-calls-invest-data-spaces-tourism-and-cultural-heritage, accessed on 10/05/2024
9. EGI Check-in - a proxy service that operates as a central hub to connect federated Identity Providers with EGI service providers, https://www.egi.eu/service/check-in, accessed on 02/05/2024
10. EUreka3D - European Union's REKonstructed content in 3D, https://eureka3d.eu, accessed on 02/05/2024
11. European Commission opens calls to invest in data spaces for cultural heritage and tourism, https://www.ne-mo.org/news-events/article/european-commission-opens-calls-to-invest-in-data-spaces-for-cultural-heritage-and-tourism, accessed on 10/05/2024
12. Europeana: discover Europe's digital cultural heritage, https://www.europeana.eu, accessed on 10/05/2023
13. Fernie, K., Blümel, I., Corns, A., Giulio, R., Ioannides, M., Niccolucci, F., Beck, J., Mathys, A., Rossi, V., Vastenhoud, C., et al.: 3d content in europeana task force. The Hague (2020)
14. Google arts & culture, https://artsandculture.google.com/project/openheritage, accessed on 10/05/2024
15. Ioannides, M., et al.: Study on quality in 3D digitisation of tangible cultural heritage: mapping parameters, formats, standards, benchmarks, methodologies, and guidelines. Tech. Rep. 2020/654, VIGIE (2020)
16. Sketchfab: the leading platform for 3D & AR on the web, https://sketchfab.com, accessed on 10/05/2023
17. Kompakkt'cause the world is multidimensional, https://kompakkt.de, accessed on 10/05/2023

18. Münster, S., Apollonio, F.I., Bluemel, I., Fallavollita, F., Foschi, R., Grellert, M., Ioannides, M., Jahn, P.H., Kurdiovsky, R., Kuroczyński, P., Lutteroth, J.E., Messemer, H., Schelbert, G.: Infrastructure, pp. 189–198. Springer Nature Switzerland, Cham (2024). https://doi.org/10.1007/978-3-031-43363-4_9
19. NFDI: Data as a common good for excellent research, organized by science in Germany, `https://www.nfdi.de`, accessed on 10/05/2024
20. Nishanbaev, I.: A web repository for geo-located 3D digital cultural heritage models. Digital Applications in Archaeology and Cultural Heritage p. e00139 (2020)
21. OpenFaaS — Serverless Functions Made Simple, `https://www.openfaas.com/`, accessed on 10/05/2024
22. Orzechowski, M., Wrzeszcz, M., Kryza, B., Łukasz Dutka, Słota, R.G., Kitowski, J.: Indexing legacy data-sets for global access and processing in multi-cloud environments. Future Generation Computer Systems pp. 150–159 (2023). https://doi.org/https://doi.org/10.1016/j.future.2023.05.024
23. Smithsonian Institution - the world's largest museum, education, and research complex, `https://www.si.edu`, accessed on 10/05/2024
24. Svoboda, T., Raček, T., Handl, J., Sabo, J., Rošinec, A., Łukasz Opioła, Jesionek, W., Ešner, M., Pernisová, M., Valasevich, N.M., Křenek, A., Svobodová, R.: Onedata4sci: Life science data management solution based on onedata. arXiv:2311.16712 (2023). https://doi.org/10.48550/arXiv.2311.16712