The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume XLVIII-M-9-2025
30th CIPA Symposium “Heritage Conservation from Bits:
From Digital Documentation to Data-driven Heritage Conservation”, 25–29 August 2025, Seoul, Republic of Korea

Quality Certification in Data Acquisition for Cultural Heritage: The Power of Paradata for High Quality Digital 3D Cultural Heritage Assets

Marinos Ioannides [1], Drew Baker [1], Athos Agapiou [2], Petros Siegkas [3], Anthony Cassar [4]

[1] Department of Electrical Engineering, Computer Engineering and Informatics. Cyprus University of Technology - marinos.ioannides@cut.ac.cy, drew.baker@cut.ac.cy

[2] Faculty of Engineering and Technology, Department of Civil Engineering and Geomatics. Cyprus University of Technology Cyprus - athos.agapiou@cut.ac.cy

[3] Faculty of Engineering and Technology, Department of Mechanical Engineering and Materials Science and Engineering – petros.siegkas@cut.ac.cy

[4] Head of Technology and Experience Development - Heritage Malta, Malta
digitisation@gov.mt

Keywords: Digital Cultural Heritage, Digital Acquisition, EU VIGIE 2020/654, Complexity Analysis, Quality Analysis, Paradata

Abstract

Published in 2022, the European Commission’s study VIGIE 2020/654: Study on Quality in 3D Digitisation of Tangible Cultural Heritage provided a comprehensive overview of the state-of-the-art in Cultural Heritage (CH) digitisation, proposing detailed guidelines toward a standardised framework for evaluating digitisation projects. The Study introduced a key distinction between the Complexity of the data acquisition methodology and the Quality of the expected results. It argued that Quality can only be meaningfully assessed by understanding the preconditions - namely, the Complexity - surrounding data acquisition (DAQ). By managing risks, faithfully documenting paradata (i.e., the processes and methods used to create the digital record), and reducing DAQ complexity, confidence in the digitisation outcomes is increased and better reflected in the overall Quality of results.

In the rapidly evolving landscape of digital technologies and their application to the CH domain, alongside the growing need to monitor monuments and sites affected by natural and human-induced hazards and interventions, this article reflects on three years of real-world implementation. It reviews the lessons learnt, identifies existing gaps, and explores emerging technologies to enhance and adapt the Study to the current and future needs of the sector. These insights are essential to ensure the continued relevance and effective adoption of the Study’s recommendations.

1. Introduction

The Study results have been implemented across several EU projects, including the H2020 ERA Chair Mnemosyne, Digital-Europe EUreka3D, Horizon-Europe ENIGMA, Horizon-Europe HERITALISE, Horizon-Europe TRIQUETRA, and Digital-Europe EUreka3D-XR. Furthermore, the association of paradata with a high-quality documentation of digital assets – enabling their use and reuse – has been successfully integrated into the Europeana Data Model (Europeana Foundation 2017).

While there is a general consensus that the (London Charter 2005) recommendations and the guidelines presented in the Study mark a significant step forward for 3D Cultural Heritage DAQ, challenges remain. The high standards expected, along with misunderstandings around terminology and rationale, which are common in multidisciplinary environments, have arguably hindered broader adoption.

If Cultural Heritage documentation is to support a meaningful certification of Quality, the following key issues must be addressed.

1.1 Complexity in Relation to Quality of an Asset

The Study offers a strategy for understanding the Complexity of the 2D/3D digitisation task, focusing on methodology and process, not just the geometric complexity of the object itself.

Figure 1 shows an example of a complexity assessment radial chart (see VIGIE 2020/654 2022, p. 62) for the digitisation of the fishing trawler Lambousa, digitised as part of the EUreaka3D project, and representing Cyprus in Europeana’s TwinIT! Campaign 2023 is an example of the re-use of 3D cultural heritage assets.

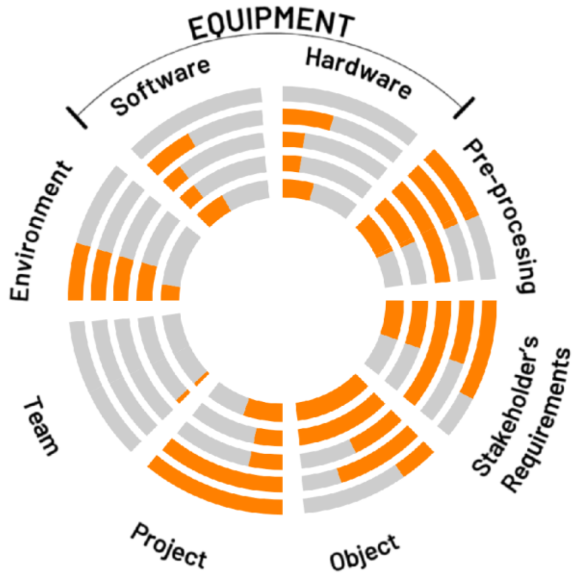

Figure 1. Complexity Assessment for digitisation of the Fishing Trawler *Lambousa*.

Full documentation for this case study can be found in *EUreka3D Report on Paradata for the documentation of The Trawler Lambousa* (EUreka3D 2024)

Reducing the Complexity of DAQ does not diminish the richness or quality of the resulting digital asset. Rather, low Complexity indicates that potential risks to successful DAQ have been anticipated and mitigated. Conversely, high Complexity signals a greater risk of compromised data acquisition and outcomes. Another important challenge is how to reduce the DAQ costs by keeping them constant and/or increasing the quality of the final result.

1.2 Quality as a Panacea

A common misunderstanding is that the Study's Quality metric directly corresponds to the visual or technical quality of the final, post-processed asset. In fact, the Study defines the final Quality of the results in terms of how thoroughly the DAQ process meets the documented expectations and reduces Complexity. This metric certifies the process, not only the product, thus ensuring high-confidence resources for multidisciplinary reuse. (Ioannides et al., 2024; Huvila et al., 2024).

Figure 2 provides an example of the companion Quality Assessment radial chart for the fishing trawler Lambousa presented in Figure 1 (see VIGIE 2020/654 2022 p. 68)

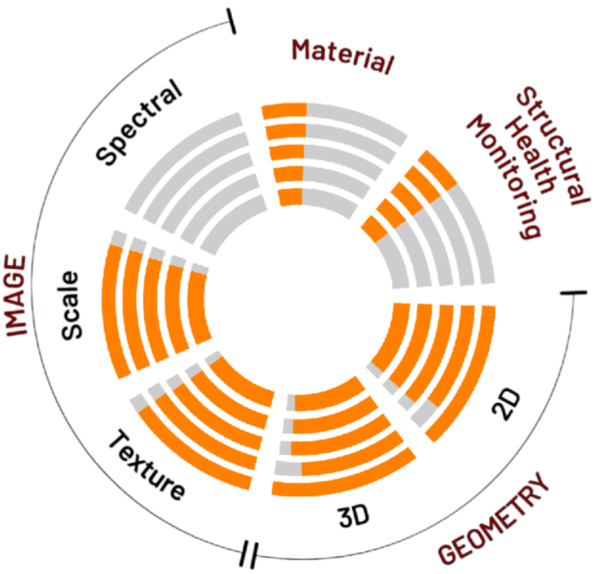

Figure 2. Quality Assessment for digitisation of the Fishing Trawler *Lambousa*.

1.3 The Burden of Documentation

As we move further into the digital era, global challenges such as climate change, armed conflict, and mass displacement are increasing the urgency for Cultural Heritage digitisation. This is reinforced by initiatives like the Commission Recommendation (EU) 2021/1970 (European Commission, 2021) and UNESCO's Dive into Heritage (2023). Follow-up research to the Study aims to support the community by advocating for monitoring monuments and sites, as well as to standardise, structure, and parameterise metadata and paradata. This approach will enable certification of methodological quality, long-term data preservation, effective asset monitoring, and compliance with the FAIR (Wilkinson et al., 2016) and CARE (Carroll et al., 2020) principles.

1.4 The Holistic Approach

We advocate a holistic approach to DAQ, aligned with the criteria set out in the Study, as a basis for reliable Complexity and Quality assessment. Holism integrates all relevant components and their relationships, enabling better identification and interpretation of gaps, needs, and interdependencies in the DAQ process. It also helps determine the required level of effort for a given project. Importantly, this approach is bound by Complexity, not the misconception that all possible data points must be captured. Understanding when, where, and how to stop the DAQ process is key to delivering meaningful and efficient Complexity and Quality documentation.

1.5 Temporal Comparison

The Study outlines when and how to assess Complexity and Quality during the DAQ process using adjustable parameters to ensure confidence in results. While this supports a single DAQ instance, there is growing recognition of the need for temporal comparison and/or real-time monitoring, tracking changes in cultural heritage objects over time. This is particularly crucial for monuments and sites requiring ongoing inspection for conservation, risk mitigation, and sustainable management.

The updated Study will address how to link multiple datasets via their metadata and paradata, allowing the reconciliation of changes in methodology and improvements over time. This will strengthen the sustainability, continuity, and long-term value of digital assets, especially for the sectors of conservation, 3D fabrication, and valorisation of cultural heritage. Establishing meaningful connections and weighted assessments across DAQs of the same asset is vital for ensuring digital longevity and resilience.

2. Towards and Updated Version of the Study

In this paper, we propose a new, updated version of the Study's recommendation, reflecting its practical use, clarifying its terms and incorporating technological advances since its original conception. Therefore, this updated version aims not only to integrate recent technological advancements, including satellite imagery, airborne sensors, low-altitude, and situ monitoring systems within its scope, but also to enhance diachronic monitoring, assessment efforts, initiatives like the European Collaborative Cloud for Cultural Heritage and new long-term data preservation standards.

Through this updated version of the Study, we are sure to engage a wider audience in the multidisciplinary Cultural Heritage community, and through the DAQ certification to build confidence in the (re)use of enriched, high-quality digital assets with a citable provenance for research, education and development in the CH domain.. Figure 3 (Overview from the Eureka3D project) illustrates the integration of the VIGIE2020/654 updated version of paradata in the digital 2D/3D documentation pipeline.

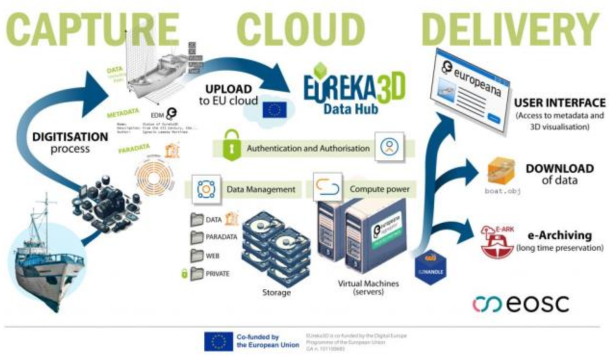

Figure 3. Overview from the Eureka3D project.

The multidisciplinary user needs encompass the diverse and intersecting requirements of both experts and non-experts from fields such as history, archaeology, computer science, museology, chemistry, civil and material engineering, surveying, and education. These users engage with digital heritage resources for a variety of purposes - including research, preservation, education, and public engagement - necessitating tailored access to data, interpretative tools, and user interfaces. The user landscape further includes policymakers, architects, electrical and mechanical engineers, geologists, artists, creative industry specialists, students, and scholars, each bringing distinct expectations and use cases.

For instance, policymakers may seek evidence-based insights to inform cultural regulations; architects and engineers require accurate spatial and material data for restoration or adaptive reuse projects; creative professionals look for authentic content to inspire innovation; and students and scholars depend on well-documented, trustworthy sources for learning and academic research. A common and critical requirement across all these user groups is access to precise, certified, and high-quality information about digitised cultural objects. Users expect rich metadata and comprehensive documentation, as well as assurances regarding the authenticity, accuracy, and reliability of the data they work with.

To meet these expectations, it is essential to implement rigorously defined methodologies and standardised processes, engage highly qualified professionals, and utilise calibrated, high-precision equipment under controlled environmental conditions. Additionally, the physical state of the object - including surface condition, material composition, and environmental influences - must be thoroughly recorded and factored into the digitisation process to ensure contextual accuracy. Answering fundamental questions -Who did What, When, Where, Why, How, under Which conditions, and in relation to which materials and surface states - must be explicitly documented and certified through Paradata.

This rigorous and transparent approach forms the foundation of the VIGIE2020/654 Study and underpins the introduction of Paradata: The Quality Certificate for Information and Data. This concept formalises and guarantees the trustworthiness, quality, provenance, and contextual integrity of cultural heritage data in digital environments.

The significance of such certification extends beyond academic research. As the DCH domain matures, there is growing recognition of the economic potential of digital heritage assets, particularly in the context of reuse within a circular economy model. While this aligns with the original vision of Paradata - to help realise a return on investment for cultural heritage research - it also shifts the value chain toward a producer-consumer model less aligned with the knowledge economy and more with traditional physical product economies.

Unlike physical objects, which have finite lifecycles, digital objects function as living repositories of knowledge. Their reuse, particularly in the creative industries and video game sectors, represents a direct output of digitisation efforts and is increasingly seen by funding bodies as a key area of impact. However, the concepts of “use” and “reuse” must be decoupled: the former pertains to the original intent and audience of a digitisation project, while the latter demands a new context and purpose, necessitating clarity, access, and standardisation.
Three conditions must be met for successful reuse:

1. Access to the digital asset;
2. Standardisation of formats, protocols, and documentation;
3. Comprehensibility of what the asset represents and how it was created.

These are embodied in the FAIR principles - Findability, Accessibility, Interoperability, and Reuse - and directly supported by the proposed certification process. The VIGIE2020/654 Study introduces a framework for quality management through the parameterisation of paradata, enabling both quality assurance and quality control of the digitisation process via two central concepts: Complexity and Quality.

- Complexity assessment focuses on the process of digital acquisition, not the inherent complexity of the object. It involves 40 parameters that define stakeholder requirements, risks, and operational contexts (such as HR, object description, project definition, etc.)
- Quality assessment evaluates the outcome of the digitisation process using 35 parameters, ensuring alignment with project specifications and use-case requirements (such as 2D, 3D, Spectral, Scale, Material, Structural Changes Health monitoring, etc.).

Together, they offer an algorithmic foundation to determine whether a digital asset is fit for purpose. What constitutes “high quality” (e.g., geometric precision or visual fidelity) is use-case dependent - what suffices for 3D printing may not meet the standards required for immersive VR applications.

The dual function of the certification is thus:

1. To certify that a quality assurance plan has been implemented for a given use-case;
2. To provide a minimum standard of transparency and reusability via documented Paradata and adherence to protocols.

While the Study provides a high-level theoretical model, its utility must extend beyond the academic sphere. To facilitate this, an updated version of the Study will include best practices, case studies, and toolkits aimed at guiding smaller institutions, community groups, and cultural stakeholders in the recording and interpretation of Paradata. This is a key component of the Memory Twin Initiative, which envisions engaging a broader multidisciplinary community in the preservation and reuse of cultural heritage.

This vision repositions certification as a catalyst for inclusive participation, enabling new contributors to maintain high standards while acknowledging their role in the digital preservation of memory. Certification reframes the re-user from a passive consumer to an active participant in a dynamic ecosystem that promotes innovation, education, and societal transformation.

However, this transformation is not without challenges. Cultural heritage is inherently dynamic - subject to physical, environmental, and social change - and the need for long-term monitoring and digital preservation is increasingly urgent. While the original Study anticipated a temporal dimension to digitisation, the proposed update seeks to integrate this more deeply, especially in response to global issues like climate change. Demonstrating its impact on heritage sites, which often carry deep cultural and identity significance, can drive public awareness and policy action more effectively than abstract data.

As part of this evolution, we are collaborating with international bodies - including Europeana, the International Image Interoperability Framework (IIIF), the EU eArchiving Initiative, and ISO committees - to develop interoperable schemas, formats, and protocols that support sustainable archiving and reuse. Additionally, a companion Paradata Recording Tool is being developed to simplify documentation, empowering smaller institutions and community-led initiatives to contribute meaningfully to global digitisation efforts.

Ultimately, this evolving framework supports a future in which digital cultural heritage is not a static archive but a resilient, interconnected, and collaboratively enriched ecosystem. Through quality certification and robust paradata, the field is laying the foundation for sustained scholarly value, creative reuse, and cultural continuity across generations.

To support his endeavour, we are developing a companion tool to make the recording of paradata for digital acquisition of cultural heritage material a more straightforward process enabling (and encouraging) smaller knowledge intuitions, heritage owners/stakeholders and community initiatives to undertake high quality documentation of underutilised heritage resources. This is a key objective in developing the Memory Twin Initiative a concept launched at EuroMed2024 and which seeks to engage a wider multidisciplinary community with cultural heritage through collaboration and cooperation.

This approach is similar to that originally envisioned for use of paradata but crucially expands the potential catchment area for new data contributions outside of the academy. It is here that the certification of quality plays another important role enabling new contributors to both maintain high-quality data standards in their contributions but also acknowledging their contribution to the digitisation effort and role in preserving cultural memory. Certification therefore can reframe the re-user role from passive consumer of project output to an active participant in the ongoing role of cultural heritage as a catalyst for innovation, education and positive societal change.

However, this move to a more dynamic approach to the perception of heritage as a living entity with a life beyond a project timeframe is not without its challenges. This should not be surprising as cultural heritage is inherently dynamic; it is part of the environment in which it exists and subject to the same pressures as its physical context.

There is an increasing need for digital monitoring of cultural heritage both to help in the conservation and protection of the physical asset but also to draw attention to the changes within the context in which they exist, for example we may discuss the climate crisis in intellectual and abstract terms but showing its effects on cultural heritage something which is usually iconic or around which there is an investment in identity and memory has more impact and meaning.

While the Study anticipated the need for a temporal facet to the digitisation of cultural heritage this main focus was to establish a starting point from new research could be developed. The proposed update to the Study recognises the increased demand for long-term monitoring requirements and is investigating how this can be integrated into the overall digitisation process and data preservation. Again, this places significant emphasis on the need to engage the Digital Cultural Heritage community with the long-term preservation and archiving of data sets and the urgency to develop standards and protocols that will enable this.

2.1 Extension of the Complexity Radar Chart for the inclusion of In-Situ and Space Monitoring

In the context of preserving Cultural Heritage (CH), the integration of advanced monitoring technologies is critical for assessing risks, mitigating damage, and ensuring sustainable conservation. Three primary approaches dominate the current landscape: **In-Situ sensors**, **UAVs/drones**, and **satellite-based space monitoring**. Each offers distinct capabilities and complexities across technical, operational, and strategic dimensions. The UAVs/drones option can be fully integrated with the previously available version of the Study published in 2022. The following presents a structured evaluation of these systems based on key complexity parameters.

2.1.1 Technical Complexity

- In-situ sensors provide high-resolution, real-time data on structural conditions (e.g., temperature, humidity, vibration) at a micro level. They require installation, calibration, and maintenance but offer unparalleled precision for localised monitoring.
- UAVs offer high-resolution photogrammetry and LiDAR data for site-scale 3D documentation but are limited by battery life, weather conditions, and regulatory constraints.
- Satellite monitoring supports wide-area surveillance, enabling environmental risk assessment, ground deformation tracking (e.g., using InSAR), and long-term landscape analysis. These systems are highly complex but require no physical access to the site.

2.1.2 Deployment Complexity

- In-situ systems demand on-site access and infrastructure for installation and operation, making them less feasible in inaccessible or conflict-prone areas.
- UAVs require field teams and flight permits but offer flexibility in data collection when access is possible.
- Satellites, in contrast, offer remote access to any site globally, eliminating deployment barriers and enabling regular, scalable monitoring.

2.1.3 Data Management Complexity

- In-situ sensors generate moderate data volumes, usually transmitted via local networks and analysed locally or in the cloud.
- UAVs produce large image and point cloud datasets requiring extensive post-processing.
- Satellite data, often available through open platforms (e.g., Copernicus, ESA, NASA), involve high computational processing for imagery, classification, and time-series analysis, but support seamless integration with GIS and remote sensing platforms.

2.1.4 Cost and Sustainability

- In-situ solutions have relatively low setup costs per site but higher long-term maintenance needs.
- UAVs require moderate capital investment and are cost-effective for single or periodic site visits.
- Satellite platforms involve significant infrastructure costs at the development level but offer free or low-cost data access and excellent long-term sustainability across broad regions.

2.1.5 Suitability for Cultural Heritage Monitoring

- In-situ sensors are ideal for detailed monitoring of structural health, microclimate, and early warnings of material degradation.
- UAVs are optimal for accurate 3D mapping, damage documentation, and visual inspection of hard-to-reach areas.
- Satellite monitoring excels at macro-scale observation, detecting land movement, vegetation encroachment, and environmental hazards, making it indispensable for regional cultural heritage risk management and inaccessible heritage sites.

An integrated, hybrid approach - leveraging the strengths of all three systems - provides the most robust and scalable solution for monitoring cultural heritage in dynamic environments.

In Figure 4 we illustrate the extension of the DAQ Complexity by integrating fully the Satellite monitoring and earth observation, which provides a powerful, scalable means of monitoring the environmental conditions and risks affecting cultural heritage sites worldwide.

Key observation domains include **land and vegetation dynamics**, **water and hydrology**, **atmospheric conditions**, and **climate-related disasters**. These are monitored using a range of specialised satellite sensor technologies such as:

1. **Optical sensors** (panchromatic, RGB) support high-resolution visual monitoring of land use, site changes, and surface conditions.
2. **Multispectral and Hyperspectral sensors** allow for detailed analysis of vegetation health, material composition, and surface degradation.
3. **Thermal Infrared (TIR) sensors** track surface temperature variations, aiding in the detection of thermal stress and moisture retention around heritage structures.
4. **Synthetic Aperture Radar (SAR)** enables all-weather, day-and-night monitoring of ground deformation, subsidence, and flood events.
5. **Spaceborne LiDAR** captures elevation and terrain data useful for 3D modelling and detecting structural shifts.
6. **Passive Microwave sensors** monitor soil moisture and vegetation water content, supporting assessments of drought and hydrological stress.
7. **Radar Altimetry** measures surface elevation changes in coastal or water-adjacent heritage zones.

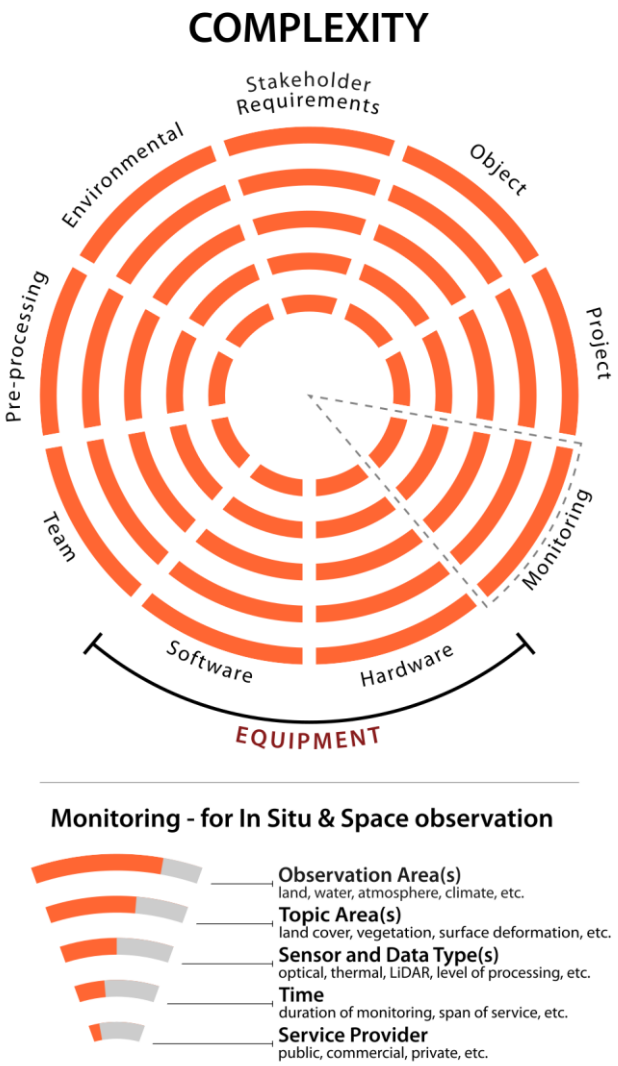

Figure 4. Data Acquisition Complexity Assessment integrating dynamic monitoring.

The combination of diverse satellite sensor types - ranging from optical and thermal to radar, hyperspectral, and altimetric - together with their long-term continuity and global coverage, provides a comprehensive, multi-dimensional understanding of the environmental pressures impacting cultural heritage sites and landscapes. These capabilities are made possible through a growing network of international Earth Observation programs and service providers, including Copernicus (EU), the European Space Agency (ESA), the National Aeronautics and Space Administration (NASA), and other national and commercial platforms, which offer open-access, high-quality datasets and specialised services that support monitoring of land use, vegetation, hydrology, atmospheric pollutants, ground deformation, and climate-related hazards. The integration of these satellite-based data streams into cultural heritage conservation workflows enables advanced early warning systems, in-depth risk assessments, and data-informed decision-making. This is especially critical in regions affected by climate change, rapid urbanisation, or limited physical access due to conflict or remoteness.

By leveraging the technological infrastructure and analytical tools provided by global Earth Observation initiatives, satellite monitoring becomes a cornerstone of modern heritage preservation - enhancing the resilience, visibility, and sustainability of cultural assets worldwide.

3. Conclusion

The extension of the VIGIE2020/654 Study to include Monitoring for In-Situ and Space Observation in Cultural Heritage represents a strategic leap forward in heritage preservation. By incorporating advanced sensing technologies, the initiative significantly improves the precision and consistency of data acquisition. This, in turn, enhances the quality and reliability of both observational data and derived insights. A key benefit is the enrichment of Paradata, providing greater transparency and reproducibility in heritage documentation and analysis. The integration of in-situ and satellite-based monitoring supports a proactive approach to risk detection and conservation planning. With richer datasets and improved data quality, stakeholders are better equipped to make informed, timely decisions. The initiative bridges technological innovation with cultural responsibility, strengthening the infrastructure for sustainable preservation. It promotes interdisciplinary collaboration and knowledge sharing. Ultimately, this extension redefines heritage monitoring as a data-driven, future-ready endeavour. It lays the groundwork for more resilient, transparent, and intelligent cultural heritage management.

Acknowledgements

The authors acknowledge the funding and support of the following EU projects in the continuing development of this work:

- **ERA Chair Mnemosyne** project has received funding from the European Union's H2020 Framework. Programme for research and innovation under Grant agreement no. 810857
- **EUreka3D** project has received funding from the European Union's Digital-Europe program under Grant agreement no. 101100685
- **EUreka3D-XR** project has received funding from the European Union's Digital-Europe program under Grant agreement no. 101174054
- **EU VIGIE2020/654** Study at the request of and financed by the European Commission, Directorate-General of Communications Networks, Content & Technology VIGIE 2020/654.
- **HERITALISE** project has received funding from the European Union's Horizon-Europe Framework. Programme for research and innovation under Grant agreement no. 101158081

References

Carroll, S, et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal, 19: XX, pp. 1–12*. DOI: 10.1038/s41597-021-00892-0

ENIGMA 2023, Horizon Europe: Endorsing Safeguarding, Protection, and Provenance Management of Cultural Heritage. https://eu-enigma.eu/ (1 March 2025)

EUreka3D 2023, Digital Europe: European Union's REKonstructed content in 3D. https://eureka3d.eu/ (1 March 2025)

EUreka3D 2024, EUreka3D Report on Paradata for the documentation of The Trawler Lambousa https://elambousa.eu/Paradata.pdf (1 March 2025)

EUreka3D-XR, 2025 Digital Europe: European Union's REKonstructed content in 3D to produce XR experiences (1 March 2025)

European Commission 2021 *Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage*, http://data.europa.eu/eli/reco/2021/1970/oj (1 March 2025)

Europeana Foundation 2017, Definition of the Europeana Data Model v5.2.8 https://pro.europeana.eu/page/edm-documentation (1 March 2025)

HERITALISE 2025, Horizon Europe: Heritage buildings and objects' digitisation & visualisation within the cloud. https://heritalise-eccch.eu/ (1 March 2025)

Huvila, I. et al 2024 *Perspectives on Paradata: Research and Practice of Documenting Process Knowledge*, Springer Cham DOI: 10.1007/978-3-031-53946-6

Ioannides, M, et al 2024 *3D Research Challenges in Cultural Heritage V: Paradata, Metadata and Data in Digitisation,* LNCS vol 15190, Springer Cham DOI: 10.1007/978-3-031-78590-0

London Charter 2005: https://www.london-charter.org/ (1 March 2025)

Mnemosyne 2019, Horizon 2020: EU ERA Chair in Digital Cultural Heritage: Mnemosyne. https://erachair-dch.eu/ (1 March 2025)

TRIQUETRA 2023, Horizon Europe Toolbox for assessing and mitigating Climate Change risks and natural hazards threatening cultural heritage. https://triquetra-project.eu/ (1 March 2025)

Europeana Twin-IT 2024: https://pro.europeana.eu/page/twin-it-3d-for-europe-s-culture (1 March 2025)

UNESCO 2023, UNESCO Dive into Heritage initiative website whc.unesco.org/en/dive-into-heritage (1 March 2025)

VIGIE 2020/654 2022 *European Commission Study VIGIE 2020/654 Study on quality in 3D digitisation of tangible cultural heritage* Publications office of the European Union DOI:10.2759/471776

Wilkinson, M., et al. 2016. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data 3, 160018* DOI: 10.1038/sdata.2016.18