IMPULSE

IMmersive digitisation: uPcycling cULtural heritage towards new reviving StratEgies

Deliverable D3.5

Report on the methodology for paradata creation and the integration of process management notations into its creation

Funded by the European Union

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Education and Culture Executive Agency (EACEA). Neither the European Union nor EACEA can be held responsible for them.

1 Document Information

Document Identification			
Status	Final	**Due Date**	31 January 2026
Version	3.1	**Submission Date**	30 January 2026
Related WP	WP3	**Document Reference**	D3.5 (21)
Related Task(s)	3.2.1, 3.2.2	**Document Type**	R
Related Deliverable(s)	D20: Guidelines on simplification of metadata based on open standards	**Dissemination Level**	PU
Lead Participant	Bruno Vandermeulen	**Lead Author**	Theodora Rontzova
Contributors	Bruno Vandermeulen, Afroditi Andreou, Anthony Cassar, Bohdan Widła, Marek Deja, Krzysztof Tomanek	**Reviewers**	Adnan Hadziselimovic, Andrew Pace

Author(s)		
First Name	**Last Name**	**Partner**
Theodora	Rontzova	KU Leuven
Bruno	Vandermeulen	KU Leuven
Marek	Deja	Jagiellonian University
Krzysztof	Tomanek	Jagiellonian University
Bohdan	Widła	Jagiellonian University
Afroditi	Andreou	Heritage Malta
Anthony	Cassar	Heritage Malta

Document History			
Version	**Date**	**Modified by**	**Modification reason**
1.0	08/01/2026	Theodora Rontzova, Bruno Vandermeulen	First draft
1.0	13/01/2026	Bruno Vandermeulen, Theodora Rontzova	Internal review

2.0	16/01/2026	Theodora Rontzova, Bruno Vandermeulen, Bohdan Widła	Internal reviewer's comments implementation
2.0	20/01/2026	Theodora Rontzova, Bruno Vandermeulen,	Second draft
2.0	25/01/2026	Bruno Vandermeulen	Internal review
2.0	27/01/2026	Adnan Hadziselimovic, Andrew Pace	Evaluation by the assigned experts
3.0	27/01/2026	Theodora Rontzova, Bruno Vandermeulen	Evaluation team's comments implementation
3.0	28/01/2026	Karolina Wawok	Evaluation by the QA Team
3.0	30/01/2026	Żaneta Żegleń Karolina Wawok Łukasz Pieczonka	SC approval and final version
3.1	01/09/2026	Bohdan Widła	Consolidated text with corrigendum from February 2026.

Quality Control		
Role	**Who (Partner short name)**	**Approval Date**
Deliverable leader	Bruno Vandermeulen	30/01/2026
Quality manager	Łukasz Pieczonka	30/01/2026
Project Coordinator	Żaneta Żegleń	29/01/206

2 Executive Summary

Deliverable D3.5 (21) presents a comprehensive methodology for the creation, management, and integration of paradata within digital cultural heritage workflows. Paradata is defined as the documentation of the processes, decisions, interpretations, and conditions behind data creation, and serves as a critical foundation for transparency, reproducibility, legal accountability, and long-term usability of digitized heritage assets. While traditional metadata describes *what* digitized or digital object is, paradata explains *how* and *why* it came to be, addressing the growing need for trustworthy, elaborate digital representations in increasingly complex immersive and multi-modal environments.

Developed within the framework of the IMPULSE Project, this deliverable responds to sector-wide challenges such as standards pluralism, interoperability gaps, inconsistent documentation practices, and the lack of tools for integrating paradata into digitization workflows. Through three cross-institutional case studies from KU Leuven Libraries (2D and multimodal imaging), Heritage Malta (3D digitization and Memory Twin framework), and the Jagiellonian University (BPMN-driven process modelling), the document synthesizes existing expertise into a unified, scalable methodology. The proposed solution is structured around a multi-dimensional, three-tier paradata model mapped across a five-stage workflow. The deliverable also introduces a mapping between paradata and Business Process Model and Notation (BPMN), enabling machine-readable, portable, and execution-ready documentation that complements existing semantic standards such as CIDOC-CRM, LIDO, PREMIS, and EDM.

Key words:

Paradata, metadata, digital cultural heritage, digitization workflows, BPMN, Memory Twin, Multimodal Imaging, Machine-readable Workflows

3 Table of Contents

4 Abbreviations and Acronyms

Abbreviation / acronym	Description
2D	Two-Dimensional
3D	Three-Dimensional
3Pi	Papyrus, Parchment, Paper Imaging Project
AAT	Art & Architecture Thesaurus (Getty)
AI	Artificial Intelligence
AI Act	Artificial Intelligence Act (EU)
AR	Augmented Reality
ARK	Archival Resource Key
BPMN	Business Process Model and Notation
CARE	Collective Benefit, Authority to Control, Responsibility, Ethics (data principles)
CAA	Civil Aviation Authority
CARARE	Archaeological/Architectural Heritage metadata standard (referenced)
CC	Creative Commons
CC0	Creative Commons Zero Public Domain Dedication
CC BY 4.0	Creative Commons Attribution 4.0
CC BY-NC-SA 4.0	Creative Commons Attribution-NonCommercial-ShareAlike 4.0
CDSM Directive	Copyright in the Digital Single Market Directive

CRM / CIDOC-CRM	CIDOC Conceptual Reference Model
CRMdig	CIDOC CRM Digital Provenance Extension
CSV	Comma-Separated Values
DC	Dublin Core
DIH	Dive Into Heritage (UNESCO)
DIP	Dissemination Information Package
DLN	Digital Lab Notebook
DNG	Digital Negative RAW format
DOI	Digital Object Identifier
EDM	Europeana Data Model
EII	Environmental and Infrastructural Information
E57	Point Cloud File Format
EU	European Union
FADGI	Federal Agencies Digital Guidelines Initiative
FAIR	Findable, Accessible, Interoperable, Reusable
FBX	Filmbox 3D Format
GCP	Ground Control Point
GDPR	General Data Protection Regulation
GeoNames	Geographic Names Database
GLAM	Galleries, Libraries, Archives, Museums
GLTF / glTF	GL Transmission Format (3D)
GSD	Ground Sample Distance

HDR	High Dynamic Range
IIIF	International Image Interoperability Framework
IMCO	IMPULSE Community of Practice
IMI	Imaging Information
IMPULSE	Immersive Digital Cultural Heritage Project
IR	Infrared
LED	Light-Emitting Diode
LiDAR	Light Detection and Ranging
LIDO	Lightweight Information Describing Objects
LOD	Linked Open Data
MARC21	Machine-Readable Cataloging Format
MR	Mixed Reality
NBMSI	Narrow Band Multispectral Imaging Standard
NER	Named Entity Recognition
NHB	National Heritage Board
OAIS	Open Archival Information System
OBJ	3D Object File Format
OCR	Optical Character Recognition
Omeka	Open-source Publishing Platform
ORCID	Open Researcher and Contributor ID
PBMN	See BPMN

PDM	Public Domain Mark
PLY	Polygon File Format
PNG	Portable Network Graphics
PREMIS	Preservation Metadata: Implementation Strategies
PROV-O	Provenance Ontology
QC	Quality Control
QA	Quality Assurance
RAW / ARW	RAW Image Format (Sony ARW)
RTI	Reflectance Transformation Imaging
RTK-GPS	Real-Time Kinematic Global Positioning System
SIP	Submission Information Package
SKOS	Simple Knowledge Organization System
STECCI	Horizon Europe cultural heritage project
SVG	Scalable Vector Graphics (contextual use)
Teneo	Preservation System (KU Leuven)
TIFF	Tagged Image File Format
TGN	Getty Thesaurus of Geographic Names
URI	Uniform Resource Identifier
USDZ	3D-Optimized Format for AR (Apple)
UV	Ultraviolet
VR	Virtual Reality
WP	Work Package

WP3	Work Package 3
XML	Extensible Markup Language
XR	Extended Reality
Zenodo	Research Data Repository

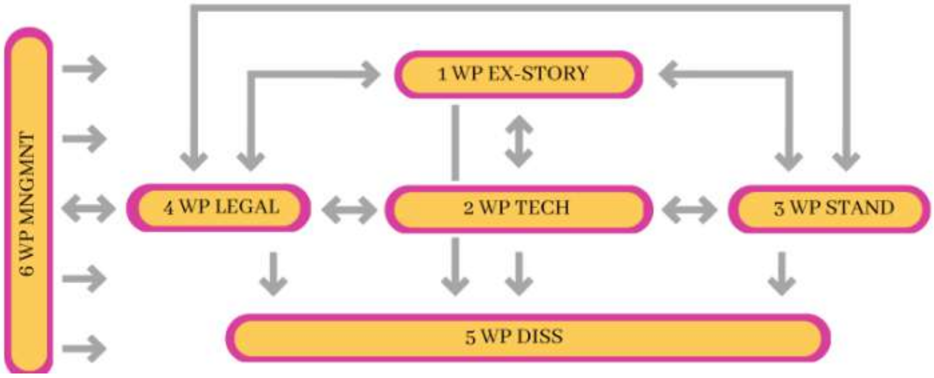
5 Introduction

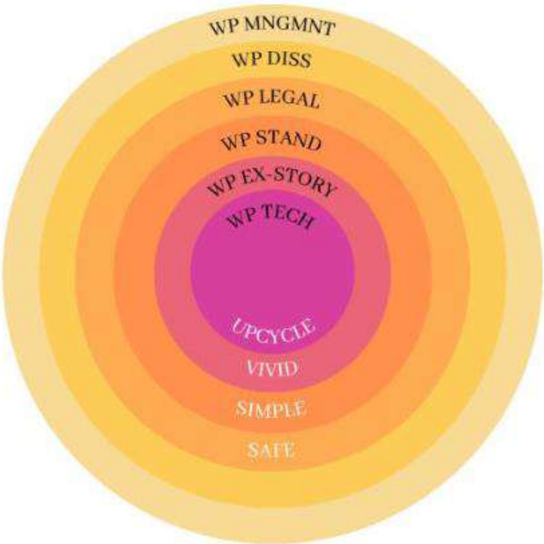
5.1. IMPULSE Project

IMPULSE emerged out of the vision of a European immersive digitization framework, driven by the forces of culture, creativity, storytelling, upcycled technology and safe, simplified standards. The project aims to synthesize innovative, multifaceted solutions and methodologies addressing the digitization and accessibility processes of the collections that make up digital cultural heritage. These initiatives aim to promote the innovative (re)use of digital cultural heritage, address challenges in platform interoperability and enhance the use of already digitized cultural heritage materials in novel contexts, such as the Metaverse and other immersive platforms. In those platforms, (whether they encompass MR, VR or AR technologies) the upcycling and appropriate reuse of digital assets remains a desideratum that our project will actively address. Additionally, IMPULSE seeks to develop pioneering standardization protocols and update the legal framework to better tackle contemporary challenges. Ultimately, the end goal of the project is to be achieved through a set of specific, measurable, achievable, realistic and timebound (SMART) specific objectives, which entail solutions that will augment the quantity and range of cultural heritage objects that are displayed through VR/XR technologies. The now easily accessible collections shall act as a powerhouse for a diverse set of demographic audiences and underrepresented communities, to empower them and help them engage with the topics and themes on display.

- Technological solutions developed within the project will enable the efficient (re)use of digitized cultural heritage content in novel contexts and immersive environments, with a focus on educational/teaching, artistic and creative dimensions, on par with the three prototypes that are being developed within the project (see IMPULSE proposal).
- Innovative standardization procedures and simplified strategies specifically targeted towards digitization processes in emerging platforms, immersive and multi-user fictional environments to achieve easily comprehended formats by deploying existing (technical) standards and metadata/paradata simplification protocols, tailored for the utilization of education, arts, and CCSI.
- Legal and organizational frameworks with detailed evaluations of risks and barriers in fields such as copyright, database rights, ownership, provenance, personal data protection and other related rights in the field of digital cultural heritage in novel environments. The lack of proper legal frameworks is an identified gap that IMPULSE aims to address, all while working within multiple national jurisdictions (namely the ones of the selected partners, i.e. Poland, Greece, Belgium, Italy, Germany and Malta) aiming to achieve harmonization.
- Connections among researchers, artists, cultural heritage practitioners and other relevant stakeholders through initiatives such as the IMPULSE Community of

Practice (referred to as IMCO), the Hackathon and the three thematic Workshops surrounding it, as well as the Acceleration & Mentoring Hub, all aiming to promote dialogue, co- creation and capacity building in immersive digitization.





IMPULSE aims to address some of the most pressing gaps in the digitization of European cultural heritage by building on existing knowledge, the capacity of its partners, and activities and networks. To achieve its stated intentions, the IMPULSE project has conceived a strategic plan which is divided into six distinct yet interconnected work packages (WPs). Each WP is indicative of the stated objectives and monitors the progress of the respective research activities and project implementation initiatives.

To conclude, the overarching objective of IMPULSE is to create innovative and comprehensive solutions that foster the digitization of cultural heritage in a standardized, findable, accessible, interoperable and reusable manner. The project is in equal parts founded on academic research and existing practice, and cons frequently on methodologies specifically fit for each Work Package.

5.2. Objectives of the Work Package 3

Within IMPULSE, Work Package 3 (WP3) emerged out of an organic need for efficient, easy to apply standardization practices in order to facilitate data sharing across various platforms, and more specifically Multi-User Virtual Platforms.

Without standardization, institutions may face several problems hampering the sharing of data. Some institutions may be reluctant to adopt certain standards as this comes in pair with significant long-term changes on systems, digital preservation, data exchange and aggregation. Some institutions may not have elaborate IT or digital resources and are dependent on software tools that have only base standards implemented, which, in some cases, the software vendor will implement to their like with no possibility to adapt.

For well-established file formats and frameworks, standards are well-documented, widely adopted, and supported by a substantial user base. In cases where these standards do not adequately address the specific needs of researchers and professionals, it may be necessary to develop supplementary or alternative standards. However, from a sustainability perspective, this approach is not always optimal. It is essential to first examine the foundational standards before considering modifications or alternatives.

When it comes to digital transition and (immersive) digitization, the very basics shall not be underestimated.

Correct and efficient standardization practices in digitization workflows ensure that the digital representations of heritage objects are consistent, reliable, findable, and comparable, both now and in the future. They enable access to and preservation of the data, facilitate sharing and collaboration among stakeholders from different EU countries, disciplines and legal frameworks, and support informed decision-making. Without the appropriate standardization practices, data may become fragmented, inaccessible, and susceptible to loss or misinterpretation over time, compromising its value and significance. Most importantly, standardization enables the creation of trustworthy data, ensuring it can be (re)used in the broadest manner.

More specifically, standardization is important in the following areas:

- Consistency: Standardization ensures that the data is captured, processed, and stored in a consistent manner, reducing the possibility of errors or unreliable datasets.
- Interoperability: Standardization enables digital heritage data to be shared between different systems and platforms, reducing the risk of data loss, inaccessibility or incomprehension (fuzzy data). Describing the genesis of datasets (equipment, software, algorithms used) will help understand and interpret the data.

- Long-term preservation: Standardized data is more likely to be preserved and accessible in the future, ensuring its long-term availability.
- Improved accuracy: Standardization provides clear guidelines and protocols for capturing, processing, and storing data, reducing the possibility of human error, and improving accuracy.

It is no exaggeration to claim that today, the cultural heritage sector is at the forefront of standardizing practices. Within heritage institutes digitization practices have been revolutionized by the development of ISO standards such as the ISO19264 image quality standards that ensure high quality, consistent data at base level for 2D objects or, on the other end of the spectrum ISO 14721:2012 Reference Model for an Open Archival Information System (OAIS) to name a few.

Practical guidelines based on the ISO 19264 standard have been developed and implemented by national institutes both in Europe (such as the Metamorfoze Preservation Imaging Guidelines) and the US (Federal Agencies Digital Guidelines Initiative) FADGI Guidelines. Many cultural institutions have implemented these guidelines, and by building upon them have generated additional practical workflows for specific contexts, for example the Rijksmuseum manuals for photographing 2D objects material types (manuscripts) or data types (3D data, audiovisual material, compression standards for video or for image). For common data formats standards and guidelines have been developed, but there is a lack for standards deploying advanced imaging techniques such as multi-spectral imaging, multi-light reflectance imaging loop characteristics in the case of creation of multimodal datasets.

In terms of metadata, several standards exist for various metadata types, e.g. technical metadata (exif, IPTC); descriptive metadata (Dublin Core, Marc21, Mods, EAD, Cidoc/CRM, Premis, metadata schema for 3D objects) etc. depending on material types, and depending on type of collection: library, archive or museal. For data interoperability and sharing of data and to ensure the information is consistent, accurate and accessible across systems, several models have been developed: Linked Open Data, SKOS, Europeana Data Model (EDM), Schema.org or International Image Interoperability Framework (IIIF). Data preservation guidelines provide information for the long-term preservation and management of digital data, ensuring that information remains accessible and usable: OAIS (Open Archival Information System) Reference Model, ISO 14721:2012.

However, as datasets expand in size at a rapid pace, novel digitization and imaging techniques emerge, and new digital preservation strategies are constantly developed. While standardization itself provides clear frameworks, the abundance of different standards that coexist often simultaneously poses several challenges; this standards "pluralism" hampers interoperability, consistency in classification and cooperation for support and development.

The adoption of standards by cultural institutions remains a challenge. Implementing and conforming to standards can be expensive for institutes with limited budgets, as it is a process that requires specialized technical knowledge and skills.

A substantial number of institutes have implemented software tools based on proprietary technology, sometimes making it difficult to share and exchange data with other systems. There is also a resistance to change and a reluctance to share data for a variety of reasons, even more so towards novel immersive environments.

IMPULSE tries to answer these challenges through research on simplifying (meta)data sharing strategies. This move aims to lift away many barriers, as it is easier for different systems and platforms to exchange and (re)use data, reduces the cost and effort of data transformation and data sharing; and opens opportunities for innovation and experimentation, fostering new and creative reuse of digitized heritage. Emerging data sharing platforms should play an important role in streamlining data to ingest into their systems, instead of outsourcing this task to content providers. This requires not only existing metadata and technical standards; but also, paradata and other documentation that describe the context and conditions of data creation and (future) data use.

5.3. Objectives of the Task 3.2.3

Task 3.2.3 aims to develop a structured and interoperable methodology for the creation and documentation of paradata in digital cultural heritage workflows, by focusing on the description of the context, processes, decisions and conditions that shape the creation and processing of data. This task seeks to formalise paradata as a complementary and emerging layer to existing metadata, by linking descriptive information about digital assets with transparent accounts of how and why they were produced, and what decisions entailed their creation.

A core objective is to investigate and integrate object-oriented process management notations such as BPMN into paradata creation, enabling workflows and decision chains to be expressed in both human- and machine-readable forms. The methodology is designed to be cross-institutional, scalable, and adaptable to different dataset types and levels of documentation maturity, providing a foundation for future automation, interoperability, and advanced reuse of digital cultural heritage resources.

6 Introduction to Paradata

In today's data-driven world, understanding how data is created, managed, and used is essential for ensuring its reliability and meaningful reuse. The concept of *paradata,* which is defined as the information about the processes and practices behind data, offers a vital lens for examining the context in which data emerges and how it can be understood. Unlike metadata, which describes the attributes of data, paradata reveals the decisions, interpretations, and actions that shape it; especially the fact that they allow the end user to better understand and interpret the data. It captures the “how” and “why” behind data, making it possible to assess its validity and applicability. Paradata is especially important because data is never neutral. It is always embedded in specific contexts, shaped by human choices, and influenced by the tools and environments in which it is produced, but also the implicit biases, present in choosing why an object/object parts came to be digitized. Paradata fills this gap by documenting the invisible labor and interpretive work that goes into data creation. Ultimately, paradata is a cornerstone of transparency and accountability in digital heritage and virtual environments. It ensures that digital representations are not only visually compelling but also epistemically sound, enabling reuse, interoperability, and critical engagement across disciplines and generations.

The concept of paradata is not confined to a single discipline. It appears in fields as diverse as archaeology, survey research, and artificial intelligence, each with its own methods for capturing and interpreting paradata. It can take many forms: field notes, logs, annotations, or even retrospective narratives, and varies depending on the theoretical and practical frameworks in use. In the field of cultural heritage, paradata plays a crucial role in preserving not only artifacts and historical records but also the interpretive processes that shape their digital representation. When heritage objects are reconstructed, digitized, or curated, paradata documents the choices made by researchers, designers, and cultural heritage institutions. It captures the rationale behind modeling decisions, the sources consulted, and the assumptions applied during reconstruction. By making these processes visible, paradata fosters intellectual honesty, supports critical engagement, and allows future researchers to assess, challenge, or build upon existing work. In this way, it safeguards the epistemic integrity of cultural heritage in the digital age.

In the digitization of cultural heritage, paradata refers to the documentation of the processes, decisions, and methodologies involved in the creation, transformation, and preservation of digital assets, especially complex ones like 3D models or multi-modal data sets. Unlike metadata, which describes the characteristics of a digitized object (e.g., title, format, creator), paradata captures the context and rationale behind its production. This includes technical specifications, software used, environmental conditions during data capture, and the interpretative choices made by technicians and researchers. Paradata has become increasingly vital in ensuring transparency, reproducibility, and scientific credibility in digital heritage projects. As digital representations grow in fidelity and complexity, the need to understand how they were created becomes essential, not only for scholarly reuse but also for ethical and legal stewardship. Without paradata, digital models' risk being perceived as opaque or misleading, undermining their value for research, education, and public engagement.

The need for a structured approach to paradata in digital cultural heritage is an increasing demand of professionals in the field. This includes developing methods for identifying and managing it, creating systems for documenting it effectively, and understanding its temporal dimensions (whether it is planned, generated during activity, or reconstructed afterward). Such efforts are crucial for improving transparency, fostering trust, and enabling responsible data reuse. Despite its importance, working with paradata presents challenges. It is often difficult to recognize, interpret, and standardize. Many researchers and professionals operate without full awareness of the paradata they need or produce. Moreover, the effort required to document paradata is frequently undervalued, even though it plays a critical role in the integrity of data work.

In the field of digital cultural heritage, one of the most crucial elements that contribute to the efficiency and credibility of digital representations is the documentation of the workflow. This documentation is known as paradata and, along with metadata, is critical to the scientific validity and sustainability of digital heritage resources. As Ioannides et al. (2024) have pointed out, paradata is used to document the different decisions, conditions, and processes involved in the development of models. While metadata mainly refers to the characteristics of the digital object per se (for example size, resolution, file format), paradata represents the context within which a digital object has been created. Such a context may involve the conditions under which data was captured, the settings and software used, or even any interventions performed by technicians. As reproducibility and replicability of results are fundamental to scientific inquiry but also the practices of cultural heritage institutions, paradata provides the necessary openness for accomplishing these goals in the field of digital cultural heritage.
However, the use of paradata in 3D digitization initiatives remains rather variable and fragmented: sources such as SketchFab and OpenHeritage do not provide any information on paradata, and as Luengo et al. (2024) notes, many aspects of paradata creation processes are underreported, including the conditions under which data was captured or the difficulties experienced while converting the data. Such gaps in documentation can lead to major discrepancies between theory and practice that reduce the practical and theoretical value of digitized cultural heritage.

One of the major issues identified is the lack of standardized procedures for paradata documentation, as each project or platform is likely to have its own approach, resulting in fragmentation of the landscape. Cassar (2024) stresses the importance of standardization during the digitization phase of national collections, as the absence of established standards not only affects the quality and accuracy of digital objects, but also complicates their integration into larger repositories or databases, a task that is crucial for the large-scale preservation of heritage. To overcome these challenges, various suggestions have been made by researchers, including Luengo et al. (2024), which is the development of a general and unified guideline for paradata documentation. When it comes to 3D digitization projects, strong documentation is crucial. Beyond the digitization process, rights management and legal/ethical issues regarding the use of models are also important factors to use and distribute the rights legally. Despite these advances, paradata remains inconsistently applied. Many platforms and repositories lack mechanisms for integrating paradata, and over-standardization risks erasing valuable interpretive traces, while under-documentation leaves critical gaps. Best practices emphasize balancing formalization with flexibility: using metadata standards and ontologies for interoperability, while preserving informal notes, sketches, and annotations that convey context. Recording decisions through audio, video, or interactive panels, embedding paradata in annexes or linking it to external repositories, and adopting identifiers for traceability are strategies that support transparency without disrupting user experience.

7 International Case Studies on Paradata

7.1. UNESCO Dive into Heritage (DIH)

Dive into Heritage (DIH) demonstrates how immersive digital heritage experiences can integrate paradata and metadata to ensure transparency and authenticity while maintaining narrative coherence and user engagement. The initiative brings together diverse 3D assets from multiple sources, including Sketchfab, UNESCO archives, and UNOSAT reconstructions, creating a unified interactive environment that allows users to explore cultural heritage in an engaging way. However, this integration introduces a fundamental challenge: how to provide detailed documentation of the decisions, processes, and technical steps behind 3D modeling without disrupting the immersive flow that makes these experiences compelling.

The approach emphasizes that paradata must capture interpretative choices, optimization strategies, and technical constraints, such as software tools, rendering settings, and data sources, alongside metadata describing object characteristics. To achieve this, DIH proposes embedding documentation through interactive panels, annexes, or linked repositories, enabling users to access detailed information when needed without overwhelming the visual narrative. This layered strategy supports adherence to FAIR principles, while ensuring that models remain traceable and reproducible. By documenting the rationale behind design decisions and the provenance of digital assets, the platform promotes ethical stewardship and long-term preservation.

Beyond technical solutions, the work highlights the importance of standardized frameworks for paradata collection and presentation. These frameworks aim to harmonize practices across institutions and platforms, reducing fragmentation, and improving interoperability. They also address the complexity of composite scenes, where multiple models with distinct origins and processes coexist, requiring granular documentation for each element. Ultimately, DIH illustrates that immersive technologies can successfully combine aesthetic appeal with scholarly transparency, creating experiences that are both captivating for the public and reliable for research and education.

7.2. The Europeana Data Model (EDM)

The Europeana Data Model is a semantic framework designed to integrate and describe cultural heritage resources in a way that supports interoperability, transparency, and long-term reuse. It was developed to address the complexity of aggregating heterogeneous data from museums, archives, libraries, and other cultural institutions into a unified structure while preserving the richness of contextual information. EDM is based on linked data principles and uses RDF (Resource Description Framework) to represent entities and relationships, enabling cultural heritage objects to be connected within a network of meanings rather than treated as isolated digital files.

At its core, EDM introduces several key entities that organize and relate to information. ProvidedCHO (Cultural Heritage Object) represents the original item, whether physical or born-digital. WebResource describes digital representations such as images, 3D models, or audiovisual files, while Aggregation acts as a container linking the object to its digital surrogates, metadata, rights statements, and provenance. This structure allows EDM to capture multiple layers of description, including technical metadata and interpretative paradata documenting decisions and workflows. By supporting controlled vocabularies and multilingual labels, EDM ensures semantic consistency across domains and languages, which is essential for Europeana's role as a pan-European aggregator. One of EDM's strengths is its flexibility in incorporating contextual relationships. It can link objects to creators, places, periods, and events, as well as to external ontologies such as CIDOC CRM, enhancing semantic richness and interoperability. This capability is crucial for meeting FAIR principles by making data machine-readable and enabling advanced search and discovery. EDM also supports rights management and licensing information, which is vital for ethical and legal stewardship of cultural heritage assets.

Beyond metadata, EDM facilitates the inclusion of paradata, allowing institutions to document how objects were digitized, curated, and interpreted. This includes recording technical processes, software tools, and methodological choices that influence digital representations. Such transparency is critical for scholarly rigor, reproducibility, and trust. EDM's RDF-based structure enables linking to external repositories and knowledge graphs, creating a dynamic ecosystem where cultural heritage data can be enriched, validated, and reused across platforms.

Ultimately, the Europeana Data Model serves as a cornerstone for sustainable digital heritage infrastructures. It harmonizes diverse data sources, supports multilingual and cross-domain integration, and provides mechanisms for capturing both descriptive and process-oriented information. By doing so, EDM ensures that cultural heritage resources remain accessible, traceable, and meaningful for research, education, and public engagement, while fostering interoperability and innovation in the digital cultural heritage landscape.

7.3. The Digital Lab Notebook

The Digital Lab Notebook is an advanced documentation system developed by Cultural Heritage Imaging to capture the complete lifecycle of cultural heritage imaging projects in a structured, transparent, and sustainable manner. Its primary goal is to ensure that every step of the imaging process, from initial planning to final data output, is thoroughly documented, making workflows reproducible, verifiable, and aligned with best practices for scientific rigor and cultural heritage preservation. Unlike traditional lab notes, this tool integrates metadata and paradata, recording not only the technical specifications of imaging but also the interpretative decisions and contextual factors that influence outcomes.

The notebook provides a centralized environment for documenting technical details such as camera settings, lens specifications, lighting conditions, calibration targets, and environmental factors during capture. It includes customizable hardware configurations for the equipment used staff profiles etc. It also records software configurations, processing parameters, and optimization steps for creating 3D models or high-resolution images. Beyond technical data, the Digital Lab Notebook captures paradata within the rationale behind methodological choices, interpretative decisions, and adjustments made during the workflow. This includes why certain imaging techniques were selected, how constraints were managed, and which assumptions guided the reconstruction process. By embedding this information directly into the workflow, the notebook reduces the risk of losing critical context and ensures that future users can understand and replicate the process.

A key feature of the Digital Lab Notebook is its support for structured templates and modular components tailored to different imaging techniques, such as photogrammetry, Reflectance Transformation Imaging (RTI), and multispectral imaging. These templates guide users through standardized documentation practices, ensuring consistency across projects. The system also integrates controlled vocabularies and metadata schemas, making the documentation machine-readable and interoperable with external repositories. This alignment with linked data principles enables connections between raw data, processed outputs, and contextual information, creating a rich semantic network that enhances transparency and reuse.

The notebook is designed to comply with FAIR principles by structuring data in formats that support long-term preservation and sharing. It facilitates linking to external ontologies and repositories, enabling cultural heritage data to be part of broader knowledge graphs. Additionally, it incorporates rights management and attribution of information, addressing ethical and legal considerations in digital heritage work. This ensures that intellectual property and usage rights are clearly documented alongside technical and interpretative details.

Beyond its technical capabilities, the Digital Lab Notebook serves as a bridge between practical imaging workflows and scholarly transparency. It supports critical engagement by making interpretative and procedural layers visible behind digital representations, fostering trust and reproducibility. By combining structured metadata, detailed paradata, and linked data strategies, the notebook exemplifies how comprehensive documentation can transform cultural heritage imaging into a robust, credible, and sustainable practice. Ultimately, it empowers professionals to produce digital assets that are not only visually compelling but also epistemically sound, ensuring their value for research, education, and public dissemination.

8. Methodology

8.1. Foundations of the Methodology

Paradata is not an entirely unknown concept in digital cultural heritage, yet it remains a relatively specialized and emerging area within the field. However, despite its recognized importance, most institutions currently approach paradata in an ad hoc and unsystematized manner. This situation stems from several factors: the absence of universally accepted standards or schemas, limited institutional resources and awareness, and technological gaps in current publication and preservation platforms that rarely accommodate paradata fields. As a result, documentation often occurs only within the scope of specific projects or is fueled by the specific needs of the professional working on a specific project, rather than being embedded in institutional workflows. This ad-hoc approach leads to fragmented practices and to a significant risk of knowledge loss when staff turnover or project completion occurs. As digital heritage grows more complex, systematic paradata workflows are becoming increasingly necessary.

8.2. Development of a cross-institutional framework

Given that standardized paradata practices are still in their infancy and that paradata documentation often remains part of the "grey literature" within most institutions (informal, unpublished, and difficult to access) we propose an iterative methodology modeled on the practices of professionals from three IMPULSE institutions as our targeted case-studies: KU Leuven, Heritage Malta and the Jagiellonian University. KU Leuven possesses academic and professional expertise in 2D and multimodal datasets (datasets consisting of different digitization techniques and file formats on one object), demonstrated in the process flow for manuscripts within KU Leuven libraries (as an example of the documentation of 2D datasets), the daily documentation practices of professionals in the Imaging Lab and the 3Pi project (an early published example of paradata documentation). Heritage Malta is a partner in several EU projects related to the 3D digitization and safeguarding of cultural heritage and possesses valuable expertise in the documentation of paradata originating from the digitization of three-dimensional artifacts. Additionally, Heritage Malta is conducting pioneering research and work on the concept of Memory Twins. Ultimately, the Jagiellonian University provides a novel perspective that combines insights from Business Process Model and Notation (BPMN) with a methodology for paradata, which describes the context, process, and conditions surrounding the creation, collection, and use of data. The aim of this approach is to integrate object-oriented process management notations and possibly elements from other business notations into the creation of paradata and create machine-readable data that can be captured by platforms and end users and provide relevant information for interpretation.

9. IMPULSE Case Studies

9.1. Case 1: 2D and Multi-Modal Datasets: KU Leuven Libraries

9.1.1. General framework

The KU Leuven paradata policy applies to all digitization and enrichment services offered by the KU Leuven Libraries. It encompasses printed books, manuscripts, journals, newspapers, fragile heritage items, and research projects. These activities are supported by specialized facilities such as the Imaging Lab and Scan Hub, which provide camera-based and scanner-based workflows designed to ensure high-quality digital reproductions while safeguarding the physical integrity of materials. The policy is grounded in transparency and modularity, ensuring that digitization workflows can be tailored to specific project needs. International imaging standards and color calibration are strictly observed to guarantee minimal and consistent quality and reproducibility. Legal compliance is integral, covering copyright and intellectual property. Furthermore, the policy aligns with interoperability and preservation standards such as IIIF, OAIS, and PREMIS, enabling sustainable data management and long-term FAIR access. In the deliverable's appendix, we chose to include the process flow for manuscripts at KU Leuven Libraries as an exemplary strategy of handling 2D datasets. Developing clear processes and mapping the context in which these take place, helps to capture and standardise paradata creation.

Implementation Guidelines

Each project begins with a documented plan that defines objectives, legal checks, and quality control measures. Modular process flows exist for different material types, including printed books and manuscripts, and protocols specify imaging parameters and conservation requirements. Outsourcing is permitted for limited steps but requires additional documentation and quality assurance to maintain compliance with KU Leuven standards.

Required Metadata

Comprehensive metadata is essential for transparency and reproducibility. Each project must record its context, including identifiers, objectives, partners, and legal basis. Object condition and handling mode are documented alongside capture configurations such as device type, resolution, color profile, and calibration details. Process steps, including imaging, OCR, and enrichment, are logged together with preservation and publication information, specifying storage systems, access platforms, and rights statements.

QA & Risk Management

Quality assurance is embedded throughout the workflow. Imaging environments are colour-calibrated, and scanning meets FADGI-level quality standards. Conservation-safe handling and climate-controlled spaces are mandatory for fragile materials. When outsourcing occurs, vendors must provide quality control reports and meet predefined acceptance criteria to mitigate risks and ensure compliance.

Roles & Responsibilities

The Digitization & Enrichment service coordinates projects and provides expert guidance. Collaboration with metadata specialists and LIBIS ensures adherence to standards and system integration. Researchers receive support through advice, infrastructure, and virtual environments, fostering a cooperative approach to digitization and data enrichment.

Tools & Systems

The Imaging Lab employs camera-based setups in calibrated environments, including Qidenus V-cradle, copy stands and Conservation Copy Stand. The Scan Hub offers advanced scanners such as Scamax 6x1 and Book2Net Lizard, transparency sets, and OCR tools like ABBYY FineReader. Preservation systems include Rosetta, Alma Cloud, and Canto, complemented by IIIF integration for interoperability and enriched access.

Compliance & Best Practices

All decisions must be documented, and calibration logs maintained. Storage choices should be deliberate, with lifecycle events captured for preservation. Standards such as IIIF, OAIS, and PREMIS must be integrated, and rights information clarified. Outsourcing details, including vendor selection and quality outcomes, are recorded to ensure accountability and transparency.

9.1.2. The 3Pi Project

An early published instance of KU Leuven Libraries' engagement with paradata appears in the NBMSI Metadata Standard, created as part of the 3Pi (Papyrus, Parchment, Paper Imaging) project. This initiative brought together experts from the Faculty of Theology and Religious Studies, the Faculty of Arts, ESAT, and KU Leuven Libraries, and focused on cross-disciplinary research involving highly fragile written heritage: papyrus rolls, parchment manuscripts, and paper artefacts. The project combined scientific imaging, analytical techniques, conservation expertise, and the development of an efficient digital dissemination platform, all with the aim of reconstructing the origin, manufacture, and material history of documentary artefacts.

Within this framework, the KU Leuven NBMSI Metadata Standard was designed to define a structured set of metadata elements and their parameters for photographic Narrow

Band Multi-Spectral Imaging workflows. The goal is to identify which technical details must accompany each image capture so that the context of creation and the characteristics of the imaging setup are thoroughly documented. Unlike traditional EXIF or IPTC schemas, this standard concentrates on operational metadata that safeguard reproducibility, ensure transparent documentation of imaging choices, and support long-term understanding of how the images came into being. These operational datapoints (collectively conceptualized as paradata) record the circumstances and technological settings involved in the capturing process.

The standard's core objective is to establish a controlled vocabulary and a harmonized structure for all metadata fields essential in NBMSI practices. By doing so, it promotes consistent documentation across imaging sessions, enhances the comparability of datasets, and reinforces automated processing by enabling machine-readable formats. The framework also supports interoperability, ensuring that imaging data can be interpreted and reused across future projects.

A central paradata category is Environmental and Infrastructural Information (EII). This category captures the physical arrangement of the studio and the conditions during image acquisition. It includes standardized descriptors for camera orientation, allowing it to be reported either through predefined positions (e.g., vertical, horizontal) or as specific angular values. Spatial measurements such as the distance between the sensor and the object, the camera's height above the floor, or the level of the imaging table, are recorded as precise numerical values. When illumination equipment is used, its height and positioning are also documented, and the presence of a transmission panel is explicitly indicated when relevant.

Another important group of elements concerns the light sources employed during multi-spectral imaging. Each source receives a unique identification number tied to its spectral emission, retrievable via the imaging software once connected to the system. These identifiers enable researchers to determine exactly which components were used and to access technical documentation for those devices. The relative placement of light units with respect to the object is also registered, along with distances, angles of incidence, and other spatial relationships that make the light geometry reproducible.

The standard further requires detailed information on the type of illumination applied. Entries specify whether the imaging relied on reflection, transmission, or fluorescence, followed by documentation of the emitter and its spectral properties. For single-peak light sources, the peak wavelength is listed in nanometres; for broadband sources, the wavelength range is recorded. Intensity settings, whether expressed as percentages or devices specific values, are also required. Likewise, any filters mounted on the camera or on the illumination system are described with their spectral characteristics. When polarization is involved, the rotation angle of the filter is documented in degrees.

Beyond EII, other categories capture additional forms of paradata. Imaging Information (IMI) includes data about the software tools used for capture and processing, as well as image resolution specifications such as DPI. Although some systems automatically embed partial information using EXIF fields, the standard mandates explicit entries due to inconsistencies in automatic metadata generation. Data Type Information (DTI) registers file formats and any processing steps applied after capture, for instance tonal adjustments, sharpening procedures, or color space conversions. These entries are crucial for evaluating the extent to which the final image diverges from the raw data.

A key prescription of the standard is that all metadata - paradata included - must be embedded within the IPTC Image Description field in every processed file. Each element receives a unique identifier to support structured extraction and computational processing. Additional documentation stored in separate files is permitted but considered secondary, since embedding metadata within the image minimizes the risk of future detachment or data loss.

Implementation examples demonstrate how these components work together. A typical image within a multispectral stack may record camera orientation, sensor-to-object distances, identifiers for each illumination unit, the spectral characteristics of the light applied, filter configurations, and intensity levels, in addition to general information about the file and its content. Such examples reveal the standard's priority: creating a complete, unambiguous record of the imaging workflow.

Paradata plays a critical role in NBMSI because it enables researchers to reproduce imaging conditions, assess the reliability of captured data, and align NBMSI outputs with general metadata frameworks such as Dublin Core. By documenting the environment, the equipment, and the operational procedures of multispectral imaging, the KU Leuven NBMSI Metadata Standard establishes a comprehensive and durable basis for the preservation, interpretation, and reuse of multispectral data.

9.2. Case 2: 3D Datasets- The case of Heritage Malta

9.2.1. Context

Heritage Malta recognizes that true preservation must include context and meaning. To achieve this, the institution integrates metadata and paradata into its digitization workflows, ensuring transparency, reproducibility, and cultural authenticity. For Heritage Malta, paradata documents the technical and interpretive decisions behind the creation of a digital surrogate, including the equipment used, resolution chosen, conditions during data capture, and software processing methods applied. Its digitisation policy aims to standardise and certify documentation practices for both metadata and paradata. By establishing common terminologies and formats, Heritage Malta seeks to enable interoperability, long-term preservation, and meaningful reuse of digital heritage assets.

These efforts align with the FAIR principles and the CARE principles ensuring that digitization is both technically robust and ethically responsible.

Established standards like Dublin Core, CIDOC-CRM, Europeana Data Model (EDM), METS, and PBCore are highlighted as essential for organizing and preserving digital assets. However, metadata alone focuses on what the object is, not how it was digitised or interpreted.

To address this gap, paradata is introduced as critical for documenting the digitization process itself. It records decisions, methods, and technical workflows, including equipment used, resolution settings, software applied, interpretive choices, and capture conditions. While metadata preserves the object's story, paradata preserves the story of its digital creation. The absence of paradata is identified as a major barrier to quality assurance and reproducibility, leaving future users unable to assess accuracy or understand the rationale behind digitization.

Although models like CIDOC-CRM provide semantic grounding for metadata, their extension into paradata remains underdeveloped. Current challenges include the lack of consensus on what constitutes sufficient paradata and the absence of user-friendly tools for its integration. Heritage Malta, in collaboration with the UNESCO Chair in Digital Cultural Heritage at the Cyprus University of Technology, is implementing structured methodologies for paradata documentation in routine workflows and Horizon Europe-funded projects such as STECCI and Heritalise. Internal protocols now detail scanning pipelines, combining technical parameters with interpretive decisions, and promoting cross-departmental collaboration. The long-term goal is to formalize these practices at national and European levels, creating certified standards for cultural heritage data acquisition, including intangible heritage.

Implementation involves detailed internal protocols for recording scanning pipelines and fostering collaboration among curators, conservators, and technologists. Heritage Malta applies these practices in projects such as STECCI and Heritalise in which it is pioneering the Memory Twin concept, which enriches digital representations with cultural narratives and emotional resonance. This approach moves beyond object-centric digitization to include intangible heritage, oral histories, and community perspectives, creating digital records that are authentic and inclusive. The Memory Twin model required experimentation, especially when integrating different types of data (text, image, sound, 3D scans, and oral histories) into a coherent system. The complexity of connecting personal stories with physical objects and metadata exposed gaps in existing CMS structures and forced the team to develop new documentation workflows. What emerged was a more adaptable, scalable framework for managing narrative clusters, collections of material linked by a shared story or individual, rather than rigid object classifications. This approach is now being explored for future applications beyond the dockyard collection. (like Stecci and Heritalise and other upcoming projects of HM)

The STECCI project, spanning eight countries and fifteen sites, illustrates this need. Each site posed unique challenges, reinforcing the importance of systematic documentation. Digitization follows the following structured framework within this project, one that aligns with the institutional practices of IMPULSE:

1) Pre-Acquisition Planning: Define the object, justify the need, anticipate challenges.
2) Acquisition Planning: Select methods and technologies, set quality procedures.
3) Acquisition & Paradata Collection: Execute scanning and record all technical and contextual details.
4) Post-Processing: Refine models and document every step.

Memory Twin represents a conceptual and methodological advancement of the digital twin model, specifically adapted for cultural heritage. While digital twins are traditionally focused on replicating the physical structure and function of objects or systems, often through real-time data, 3D modelling, and simulation, the Memory Twin framework expands upon the former to incorporate the intangible, emotional, and social dimensions that give heritage its full meaning. Rather than serving solely as a technical representation, the Memory Twin functions as a holistic digital construct that captures not only the physical characteristics of heritage assets but also the ways they have been experienced, remembered, and interpreted within cultural and community contexts; therefore creating an assemblage of both tangible and intangible cultural heritage layers which overlap and inform each other. From Digital Twin to Memory Twin

Digital twins are virtual representations of physical assets, environments, or systems originally developed in engineering and industrial contexts; they enable monitoring, simulation, and scenario testing by linking a digital model to its real-world counterpart. However, while they excel at representing measurable and quantifiable attributes, traditional digital twins often remain limited in their capacity to capture historical context, interpretive decisions, and the intangible dimensions of cultural significance.

Conventional digital twins have their origins in industrial design and engineering, where their primary purpose is to simulate physical systems and monitor performance through real-time data. In heritage contexts, digital twins have been adapted to create highly accurate 3D models of artefacts, monuments, and sites. These digital surrogates enable research, conservation, and virtual access, especially in cases where physical heritage is at risk from environmental threats or human impact.

However, the fidelity of form achieved by digital twins often comes at the cost of contextual depth. The technological precision of a 3D scan or high-resolution image can obscure the social and emotional narratives that imbue heritage with cultural value. A monument's aesthetic features or material composition can be captured flawlessly, but without the associated stories, rituals, and lived experiences, the resulting digital object risks becoming a sterile simulation, making its history technically sound but culturally incomplete. The Memory Twin framework addresses this shortcoming by proposing a model that incorporates not only tangible data, but also intangible cultural heritage and the documentation of the digitisation process itself. It turns objects into artefacts and buildings into monuments by embedding them in the memory, experience, and meaning of the communities that created and continue to engage with them.

While digital twins focus on geometric data and technical monitoring, this process is not sufficient for the digitisation of cultural heritage assets. If digitisation is to retain the authentic identity of the Cultural Heritage (CH) asset, it needs to include additional layers of information and ensure that the digitisation acquisition and post-processing achieve the highest certified standards. Memory Twin on the other hand, introduces a critical, layered methodology for data authenticity and traceability, emphasising cultural embeddedness through narrative layers of knowledge, provenance chains, and intangible values.

It draws directly from and expands upon the Holistic Heritage Building Information Modelling (HHBIM) approach by interconnecting the technical, historical, legal, and societal dimensions of cultural artefacts in a semantically rich, transparent, and ethically grounded digital model. This enhances traditional Heritage BIM (HBIM) by addressing its limitations in geometric modelling and digital representation. HHBIM integrates historical research, material analysis, structural assessment, lifecycle management, stakeholder collaboration, and digital twins, creating a more dynamic and semantically rich approach. It aligns with EU H2020 initiatives, incorporating AI, IoT, and GIS to advance heritage documentation, conservation, and management within a multidisciplinary framework. The HBIM framework, when done accurately, consists of workflow descriptions of both current and historic data acquisitions but also importantly includes dissemination methods and 3D modelling.

The Memory Twin framework integrates metadata, meaning descriptive information about a digital object, such as file formats, spatial resolution, or software used, but also paradata, which refers to the contextual documentation of how, when, and by whom the data was generated, processed, and for what purpose. This paradata layer functions as a form of epistemological scaffolding, supporting the verifiability and accountability of digital heritage practices, and allowing future researchers to understand the interpretive decisions embedded in the modelling process (FAIR data).

Hirsch et al. (2024) highlight the need for HCI to move beyond interface design towards fostering multiperspectivity, inclusion, and sensemaking in digital heritage contexts. The Memory Twin framework builds upon moving beyond interface design towards fostering multiperspectivity, inclusion, and sensemaking in digital heritage contexts this vision by developing interdisciplinary methodologies that connect technical infrastructures with human narratives, ensuring that each digital model functions as a site of interpretation rather than a fixed representation. The project translates HCI (Human Computer Interaction) advances into cultural practice through participatory data modelling, where multiple voices, including curators, conservators, technologists, and communities, contribute to the interpretive process. This aligns with the concept of Hirsch et al.'s call for “re-contextualization” within HCI, reframing digitisation not as neutral reproduction but as a collaborative act of storytelling. Through interoperable metadata standards and participatory living labs, Memory Twin advances a humane, reflexive model of digital heritage that invites inclusion, dialogue, and critical reflection.

An example from Heritage Malta’s Maritime Museum illustrates this: a small, faded Angry Birds inflatable ball displayed in an exhibition seemed trivial until visitors learned it had been the only life-saving item a child carried during a dangerous sea crossing. This story transformed the object into a powerful symbol of identity and survival, demonstrating the essence of the Memory Twin, connecting physical artefacts with emotional and narrative depth.

To achieve this, traditional metadata and paradata practices must be extended. Conventional documentation models effectively describe physical attributes and digitization workflows but fail to capture emotional, historical, and experiential dimensions. Existing models like Holistic Heritage Building Information Modelling (HHBIM) provide a foundation for this integrated approach by embedding historical narratives into data frameworks; however, the Memory Twin requires narrative-driven metadata, incorporating oral histories, storytelling, and user interactions, and experiential paradata, which records not only how an object was digitised but also how it was historically used, interpreted, and remembered by communities. In some cases, multisensory data such as soundscapes, environmental conditions, or haptic interactions can enhance immersive engagement. While metadata records information about the object, paradata captures the human decisions and technical workflows involved in creating the digital surrogate. This includes details such as the equipment used, scanning resolutions, software settings, and interpretive decisions made during digital restoration or modelling. Paradata enhances transparency and reproducibility. It allows future users to assess the accuracy and authenticity of the digitised object and understand the epistemological lens through which it was created. This is particularly important in a time when AI-generated visualisations and reconstructions can distort or oversimplify heritage narratives if not grounded in documented methods.

The absence of paradata is one of the greatest weaknesses in current digitisation practices, especially in 3D workflows. To address this, Heritage Malta has begun developing internal protocols and advocating for standardised, certifiable frameworks for metadata and paradata. These efforts align with the FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective benefit, Authority to control, Responsibility, and Ethics) principles, which together ensure that digitisation is both technically robust and ethically grounded.

Paradata, understood as the contextual information detailing the decision-making, interpretative choices, and technical conditions underpinning the creation of digital assets, extends beyond conventional metadata to reveal the epistemic processes of heritage reconstruction. As demonstrated in these initiatives, paradata preserves not only the what of heritage but the how and why of its transmission. This approach aligns with contemporary research in digital heritage especially that of Ioannides et al. and Storeide et al., which identifies the systematic recording of paradata as essential for ensuring transparency, reproducibility, and interpretative accountability in 3D digitisation workflows.

In conclusion, The Memory Twin offers a holistic, ethical, and context-sensitive approach to the digitisation of cultural heritage. By integrating tangible data, intangible narratives, and detailed paradata, it overcomes the limitations of conventional digital twins and provides a richer, more inclusive record of the past. It aligns with critical heritage theory and at the same time responds to the ethical imperatives of digital preservation in the twenty-first century, transforming digitisation from a process of technical replication into a dynamic act of cultural remembrance. In this light, the Memory Twin framework positions digital archives as spaces of resistance against epistemicide: instead of flattening heritage into uniform data points, it preserves and exposes the interpretive plurality embedded in cultural artefacts. By treating metadata and paradata as acts of care and accountability, the project operationalises cultural restorative principles in data design, ensuring that digitisation processes remain transparent, situated, and reparative.

9.2.2. Heritage Malta Paradata Sequence

Heritage Malta has developed a framework to capture paradata:

1. Digitization Process Details

- Equipment used (e.g. camera model, scanner type, lighting setup)
- Software used (e.g. Adobe Lightroom, Capture One, photogrammetry tools)
- Settings (e.g. resolution, colour profile, file format)
- Post-processing steps (e.g. cropping, colour correction, stitching)

2. Staff Involvement

- Who digitised the object (names, roles, institutions)
- Expertise level (e.g. trained photographer, student intern)
- Decision-making rationale (e.g. why a certain angle or lighting was chosen)

3. Documentation of Choices

- Selection criteria (why this object was digitised)
- Interpretive decisions (e.g. reconstruction of missing parts, 3D modeling assumptions)
- Cultural sensitivity considerations (e.g. restrictions on sacred objects)

4. Temporal and Contextual Information

- Date and time of digitization
- Location of digitization
- Conditions during digitization (e.g. ambient light, temperature)

5. Versioning and Provenance

- Version history of digital files

Changes made over time
Links to original/raw files and edited versions

6. Workflow and Standards

Digitization protocols followed (e.g. Europeana, CARARE, Spectrum)
Quality control procedures
File naming conventions and storage practices

9.3. Case 3: Paradata and Business models: object-oriented process management – the case of Jagiellonian University

9.3.1. BPMN and GLAM Digitization

Business Process Model and Notation (BPMN) is an international standard developed by the Object Management Group to represent processes in a graphical notation that is readable across stakeholder groups, spanning early analytical drafts through to implementation and management. Its scope is explicitly oriented to end-to-end "business processes," but the standard's motivation is to bridge process design and process execution through a shared representation, including a formal model and an interchange format intended to support portability of process definitions across tools. This orientation makes BPMN directly relevant to digitization work in galleries, libraries, archives, and museums (GLAM), insofar as digitization is an institutional workflow composed of sequenced activities, role-structured responsibilities, exceptions, and coordination across organizational boundaries. BPMN is designed to communicate process knowledge to diverse audiences and supports both "as-is" representations of existing practice and "to-be" models for redesigned practice, enabling analytic road mapping and optimization prior to automation.

Executable vs. Non-Executable Processes

A key feature for GLAM applications is BPMN's explicit distinction between non-executable and executable internal processes. Non-executable private processes document behaviour at a modeler-defined level of detail and are well suited for analytical capture of digitization practice, including the identification of handoffs, bottlenecks, decision points, and accountability loci. Executable private processes, by contrast, are modelled with sufficient formal detail to be enacted according to BPMN execution semantics, supporting orchestration of worker tasks and system-mediated steps as an integrated workflow specification.

BPMN Interchange Architecture and Paradata

For paradata-oriented digitization research and platform design, BPMN's interchange architecture is especially important because the standard includes an interchange format and XML schema infrastructure intended to support import/export and tool-to-tool portability, as well as "machine consumable files" and XML-based representations of process definitions and diagram interchange. This enables digitization workflows to be preserved not only as narrative documentation but also as machine-readable artifacts (e.g., bpmn) that can be captured by workflow platforms, linked to execution logs, and delivered to end users as structured contextual information for interpretation of digitized outputs.

At the same time, BPMN should be positioned as a procedural meta-language rather than a replacement for cultural heritage ontologies and metadata standards: the BPMN standard explicitly excludes data and information modelling from its scope and notes that BPMN is not a data flow language, even though it can represent message exchange and associate data artifacts with activities. This makes BPMN complementary to established semantic and metadata frameworks such as CIDOC CRM for cultural heritage information integration and exchange, LIDO for object description interchange, METS for packaging descriptive, administrative, and structural metadata in digital library contexts, and PREMIS for preservation metadata in digital archiving systems. In an integrated framework, ontologies and metadata standards can describe "what" a digital object is, while BPMN provides a standardized representation of "how," "by whom," "under what conditions," and "in what sequence" digitization data were produced, managed, and published; thereby formalizing paradata as a portable, inspectable process layer aligned with analytical and automation use cases.

9.3.2. BPMN Integration in GLAM Workflows

BPMN can be treated as a meta-language for GLAM digitization because it expresses a process as an ordered and branching sequence of activities, events, and decisions, while remaining compatible with both documentation (as-is/to-be) and execution-oriented specifications. This framing aligns with digitization practice as a multi-step process that spans preparation, metadata creation, imaging, enrichment/processing, ingest and publishing, and post-care, rather than a single capture act, and it foregrounds the practical observation that metadata and workflow coordination are structurally central to successful digitization outcomes.

The same applicability is visible in operational protocols that describe digitization as a coordinated chain of steps across roles and systems, where work is tracked per package, transitions between steps are monitored, and quality control and issue reporting are built into the process flow.

In BPMN terms, a GLAM organization (or a tightly governed organizational unit) can be modelled as one Pool that contains its internal orchestration (a private process), while exchanges of information, packages, or requests across organizations or external platforms are represented as Message Flows between Pools. BPMN's standard semantics explicitly support private processes (executable and non-executable) and collaborations between participants, and it constrains a private process to a single Pool such that sequence flow does not cross pool boundaries, while message flow can cross to represent interactions between separate participants.

This is the structural basis for modelling digitization as (a) orchestrated internal work for recording and automation and (b) cross-boundary publication and preservation interactions (e.g., repository deposit, DOI minting, aggregation), while keeping a single machine-readable representation exportable as BPMN XML.

Roadmap for "BPMN-to-digital-heritage" modelling

This session introduces how core BPMN elements can be reinterpreted for digital heritage workflows in GLAM environments. By pairing each BPMN construct with its operational meaning the digitization process, preservation and data governance, it outlines a modelling approach that embeds essential paradata and supports documentation.

Pool (Participant).

- BPMN meaning: a participant in a collaboration and the container boundary for a private process; can be "black box" or contain internal details.
- GLAM meaning: an organization or governed unit operating one coherent workflow domain (e.g., Museum Digitization Unit; Library Imaging Lab; External Repository).
- Paradata payload: institution identifier; governance profile; primary systems-of-record; policy constraints; legal/accountability boundary.

<bpmndi:BPMNPlane id="BPMNPlane_1" bpmnElement="Collaboration">

<bpmndi:BPMNShape id="Participant" bpmnElement="Participant" isHorizontal="true">

<dc:Bounds x="160" y="80" width="600" height="250" />

<bpmndi:BPMNLabel />

</bpmndi:BPMNShape>

Co-funded by the European Union

</bpmndi:BPMNPlane>

Fig. 1. Depiction of an exemplary Interaction among Participants in BPMN.

Core BPMN Elements and GLAM Interpretations

Lane.

In BPMN, a Lane subdivides a Poll to demonstrate *who* performs *what* task within a workflow. Lanes allow the representation of the internal structure of a digitization unit by separating activities, for example to Curator, Conservator, Metadata Specialist etc.

- BPMN meaning: sub-partition within a pool used to organize and categorize activities.
- GLAM meaning: role or team partition (Curator; Conservator; Imaging Operator; Metadata Specialist; Data Steward; QA; Legal or Work Packages).
- Paradata payload: role taxonomy; training requirements; authorization scope; responsibility assignment.

Collaboration.

A Collaboration in BPMN shows how multiple independent participants (Pools) interact via message flows. This is essential within digitization workflows, as a high level of interaction is observed. A Collaboration model captures the end-to-end environment, exposing how data, physical objects, paradata, rights information, and preservation packages move between organizations.

- BPMN meaning: depiction of interactions between two or more participants (pools) connected by message flows.
- GLAM meaning: end-to-end digitization environment spanning collection holder, digitization lab, preservation repository, discovery platform.
- Paradata payload: partner registry; interaction map; data-sharing flows.


```xml
<bpmndi:BPMNPlane id="BPMNPlane_1" bpmnElement="Collaboration_1dav09m">
    <bpmndi:BPMNShape id="Partner_1" bpmnElement="Partner_1" isHorizontal="true">
      <dc:Bounds x="160" y="210" width="300" height="80" />
      <bpmndi:BPMNLabel />
    </bpmndi:BPMNShape>
    <bpmndi:BPMNShape id="Partner_2" bpmnElement="Partner_2" isHorizontal="true">
      <dc:Bounds x="160" y="80" width="300" height="70" />
      <bpmndi:BPMNLabel />
    </bpmndi:BPMNShape>
    <bpmndi:BPMNEdge id="Flow_1" bpmnElement="Flow_1">
      <di:waypoint x="230" y="150" />
      <di:waypoint x="230" y="210" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNEdge id="Flow_2" bpmnElement="Flow_2">
      <di:waypoint x="350" y="150" />
      <di:waypoint x="350" y="210" />
    </bpmndi:BPMNEdge>
    <bpmndi:BPMNEdge id="Flow_3" bpmnElement="Flow_3">
```

```
    <di:waypoint x="290" y="210" />
    <di:waypoint x="290" y="150" />
  </bpmndi:BPMNEdge>
  <bpmndi:BPMNEdge id="Flow_4" bpmnElement="Flow_4">
    <di:waypoint x="410" y="210" />
<di:waypoint x="410" y="150" />
  </bpmndi:BPMNEdge>
</bpmndi:BPMNPlane>
```

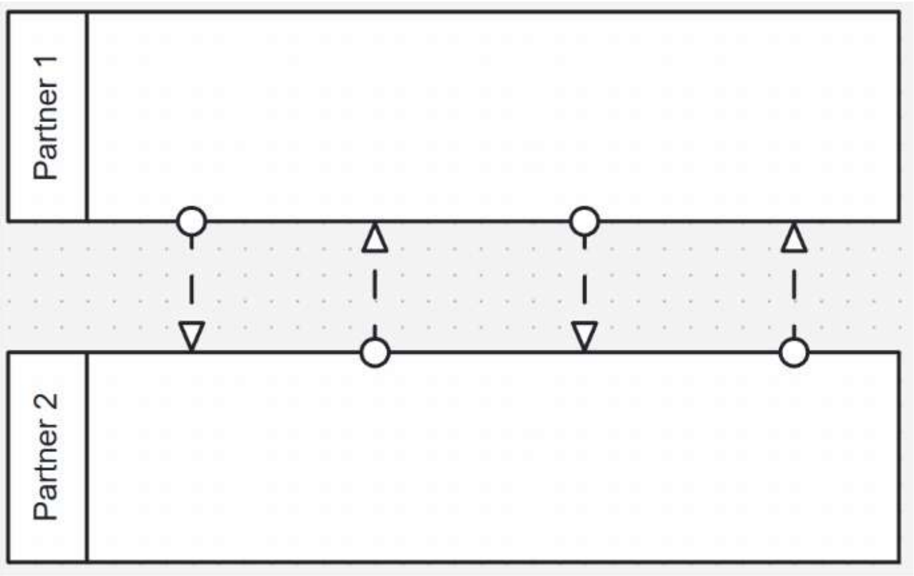


Fig. 2. Minimal Collaboration layout:
two participants interacting solely through message flows, with no internal activities displayed.

Private process (non-executable).

- BPMN meaning: modelled to document behaviour at a modeler-defined level of detail; may omit formal execution conditions.
- GLAM meaning: analytical road map of digitization practice for transparency, paradata capture, and optimization, even when no automation engine is used.

- Paradata payload: documentation completeness; evidence pointers; "as-is" traceability.

```
<bpmn:process id="AutomatedIngest_Exec" name="Automated Repository Ingest" isExecutable="true" modeler:executionPlatform="Camunda Cloud">

  <bpmn:extensionElements>

   </bpmn:extensionElements>

  <bpmn:serviceTask id="Task_Checksum" name="Calculate SHA-256" />

  </bpmn:process>
```

Private process (executable).

- BPMN meaning: modelled for execution according to BPMN semantics.
- GLAM meaning: workflow automation/orchestration specification (e.g., task lists, automated processing jobs, deposit steps) suitable for BPMN-based engines.
- Paradata payload: executable constraints; message correlation keys; runtime audit.

```
<bpmn:collaboration id="Collaboration_GLAM_Workflow">

    <bpmn:participant id="Participant_Aggregator" name="External Repository (Public Process)" processRef="Process_Public" />

     <bpmn:participant id="Participant_Digitization" name="Internal Digitization Unit (Private Executable)" processRef="Process_Executable_Digitization" />

                <bpmn:messageFlow id="Flow_Message_Transfer" name="Transfer SIP/METS" sourceRef="Activity_Package" targetRef="StartEvent_Public_Receive" />

  </bpmn:collaboration>

           <bpmn:process id="Process_Executable_Digitization" name="Digitization Workflow" isExecutable="true">

    <bpmn:startEvent id="Start_Digitization" name="Request Received" />

     <bpmn:sequenceFlow id="Flow_1" sourceRef="Start_Digitization" targetRef="Activity_Capture" />
```

Public process.

- BPMN meaning: shows only the externally visible interaction points (message exchanges) of a participant.
- GLAM meaning: "deposit/publication contract" view of a unit's workflow (what must be sent/received to collaborate), useful for repositories/aggregators and cross-institutional interoperability.
- Paradata payload: touchpoints; required metadata and packages sharing; process validation.

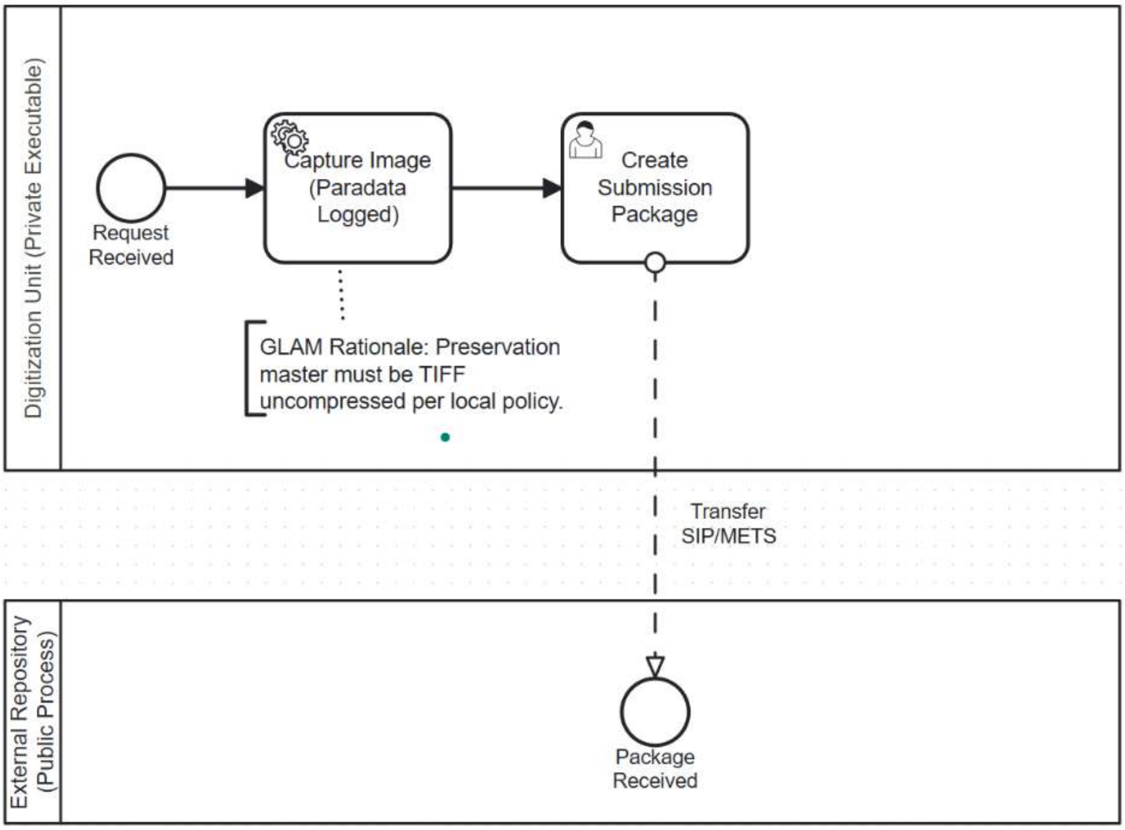


Fig. 3. Depiction of the Private and Public Process.

Sequence flow.

- BPMN meaning: ordering relation inside a process, carrying control tokens.
- GLAM meaning: the internal order of digitization steps (e.g., prepare object → capture → QA → processing → package → ingest).
- Paradata payload: step ordering, dependencies, and permissible parallelization.

Message flow.

- BPMN meaning: message exchange between participants; must not connect objects within the same pool.
- GLAM meaning: transfer of requests, packages, metadata exports, deposit confirmations across organizational/system boundaries (e.g., transfer to repository; DOI mint request; aggregator push; partner communication).
- Paradata payload: payload schema; protocol; transport mechanism; sender/receiver; timestamps; success/failure states.

Association.

- BPMN meaning: links artifacts/information to BPMN elements; does not affect control flow; may indicate direction when appropriate.
- GLAM meaning: attach evidentiary paradata to steps (checklists, calibration reports, QA logs, conservation notes) without changing the process logic.
- Paradata payload: provenance pointer (URI/ID); document type; integrity hash.

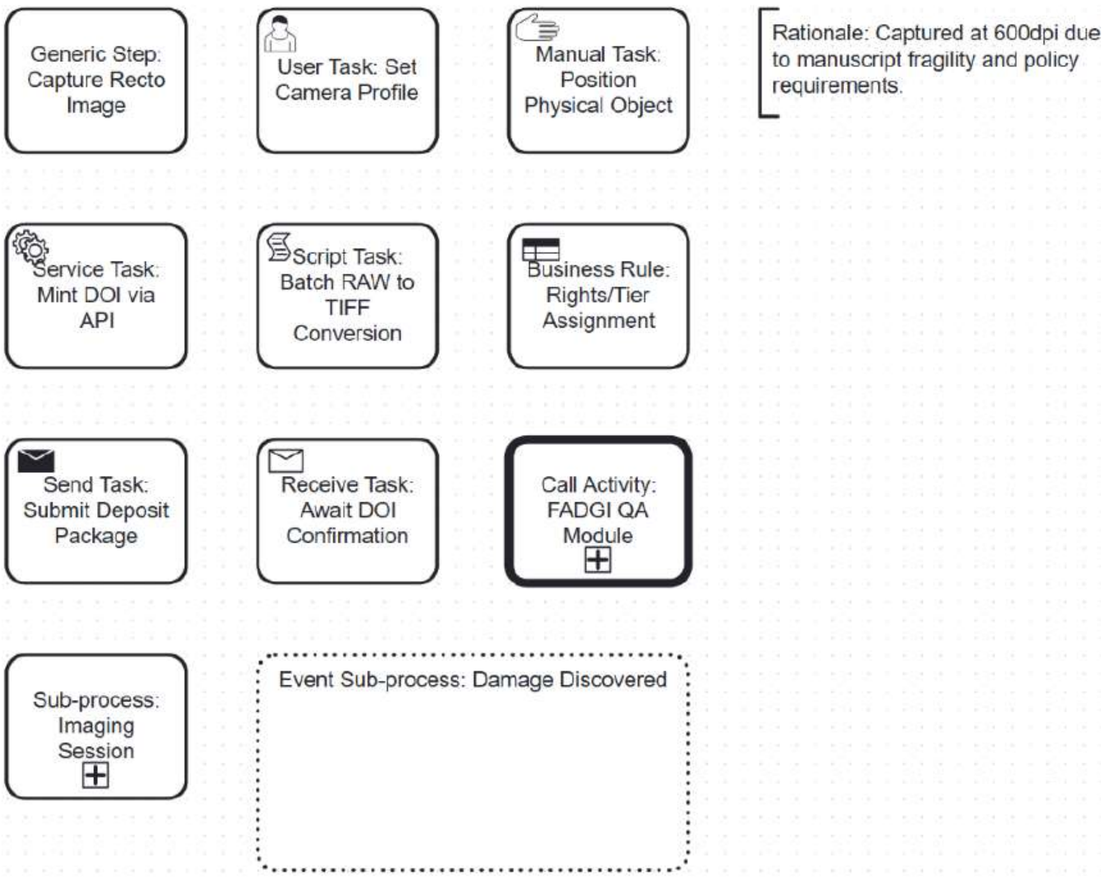


Fig. 4. Depiction of the three connector types of BPMN: Sequence Flows, Message Flows, Association.

Text annotation.

- BPMN meaning: free text information for the reader; does not affect flow.
- GLAM meaning: human-readable rationale (e.g., why a reduction was applied; why capture constraints required deviation).
- Paradata payload: decision justification; links to policy/standard clauses.

<bpmn:textAnnotation id="Annotation_Rationale"> <bpmn:text>Rationale: Captured at 600dpi due to manuscript fragility and policy requirements.</bpmn:text> </bpmn:textAnnotation>

Task (generic).

- BPMN meaning: atomic unit of work.
- GLAM meaning: one operational step (e.g., "capture recto image"; "validate checksum"; "map metadata to EDM").
- Paradata payload: performer role; tool; start/end time; inputs/outputs; environment.

<bpmn:task id="Activity_Generic" name="Generic Step: Capture Recto Image" />

User task.

- BPMN meaning: performed by a human with the assistance of a software application; lifecycle managed by a task manager/runtime.
- GLAM meaning: "operator-in-the-software-loop" steps (set camera profile in capture software; edit scan parameters; enter metadata in a controlled UI; accept QA results in a dashboard).
- Paradata payload: user ID/role; UI form/version; parameter selections; validation messages.

<bpmn:userTask id="Activity_User" name="User Task: Set Camera Profile" />

Manual task.

- BPMN meaning: performed without the aid of a process engine or application; not managed by a runtime.
- GLAM meaning: embodied craft and physical handling (position object; adjust camera angles; place calibration target; pack for transport; handle manuscripts under handling rules), consistent with protocols emphasizing trained staff and controlled handling procedures (KU Leuven Bibliotheken, 2022; OMG, 2013).

<bpmn:manualTask id="Activity_Manual" name="Manual Task: Position Physical Object" />

Service task.

- BPMN meaning: invokes an external service/application under engine control.
- GLAM meaning: automated calls to infrastructure services (checksum service; DOI minting API; repository ingest endpoint; IIIF manifest generator; thumbnail generation service).
- Paradata payload: service endpoint; request/response; retry policy; fault details.

<bpmn:serviceTask id="Activity_Service" name="Service Task: Mint DOI via API" />

Script task.

- BPMN meaning: executed by a process engine using a defined script; completes when the script completes.
- GLAM meaning: locally executed automated processing (batch RAW→TIFF conversion; photogrammetry pipeline trigger; OCR batch; derivative generation; metadata normalization script).
- Paradata payload: script identifier; version; runtime; parameters; logs; compute environment.

<bpmn:scriptTask id="Activity_Script" name="Script Task: Batch RAW to TIFF Conversion" />

Business rule task.

- BPMN meaning: provides input to a rules engine and receives calculated outputs.
- GLAM meaning: formalized policy decisions and validation logic (rights/licensing decision rules; tier assignment; selection rules; privacy redaction thresholds; format acceptance rules).
- Paradata payload: rule set ID/version; inputs (facts); outputs (decision); justification trace.

<bpmn:businessRuleTask id="Activity_Rules" name="Business Rule: Rights/Tier Assignment" />

Send task.

- BPMN meaning: sends a message and completes.
- GLAM meaning: “push” actions (submit deposit package; notify partner; send QA report; submit DOI mint request).
- Paradata payload: message schema; recipient; transmission evidence.

<bpmn:sendTask id="Activity_Send" name="Send Task: Submit Deposit Package" />

Receive task.

- BPMN meaning: waits for a message; can instantiate a process if configured; transfers message payload to data output when present.
- GLAM meaning: “await external confirmation/data” (receive partner package; receive DOI; receive rights clearance; receive validation report), including cases where a received package triggers a downstream workflow instance.
- Paradata payload: correlation key; payload capture; validation status.

```
<bpmn:receiveTask id="Activity_Receive" name="Receive Task: Await DOI Confirmation" />
```

Call activity.

- BPMN meaning: invokes a reusable process.
- GLAM meaning: reusable protocol modules (e.g., “object preparation protocol”; “FADGI QA module”; “repository deposit module”) shared across projects/collections.
- Paradata payload: invoked protocol ID/version; parameters; compliance evidence.

```
<bpmn:callActivity id="Activity_Call" name="Call Activity: FADGI QA Module" />
```

Sub-process (embedded/nested).

- BPMN meaning: encapsulated activity whose internal elements share the parent process’s data context.
- GLAM meaning: a contained stage such as “Imaging Session” or “Metadata Enrichment Stage” where many micro-steps operate on the same working dataset.
- Paradata payload: stage-level summary plus step-level logs.

```
<bpmn:subProcess id="Activity_Subprocess" name="Sub-process: Imaging Session" />
```

Event sub-process.

- BPMN meaning: sub-process triggered by an event (often for exceptions).
- GLAM meaning: exception handling modules (damage discovered; rights issue; QA failure remediation; re-capture request).
- Paradata payload: incident record; corrective action; rework scope.

```
<bpmn:subProcess id="Activity_EventSub" name="Event Sub-process: Damage Discovered" triggeredByEvent="true"> <bpmn:startEvent id="Event_Error" name="Error Detected" /> </bpmn:subProcess>
```

Fig. 5. Three BPMN activity markers showing a transaction sub-process, a multi-instance activity, and a loop activity.

Transaction sub-process [left].

- BPMN meaning: specialized sub-process supporting transaction protocol semantics including cancel/compensation patterns.
- GLAM meaning: "deposit as a controlled transaction" where partial failure requires rollback or compensating actions (withdraw package; revoke publication; replace DOI landing page).
- Paradata payload: transaction ID; rollback evidence; compensating actions.

Multi-instance activity (parallel/sequential) [central].

- BPMN meaning: repeated execution for a collection of items.
- GLAM meaning: per-object or per-file repetition (scan each folio; capture each angle; run OCR per page; generate derivatives per file).
- Paradata payload: item list; iteration count; per-item metrics.

Loop characteristics (standard loop) [right].

- BPMN meaning: looping controlled by a condition and optional maximum.
- GLAM meaning: iterate until quality threshold is met (repeat capture until focus/lighting passes; repeat metadata cleanup until validation passes).
- Paradata payload: loop condition; stopping rule; iteration evidence.

Fig. 6. Five BPMN event symbols showing a none start event, a message start event, a timer intermediate event, a message intermediate event, and an escalation intermediate event.

Start event (none) [left].

- BPMN meaning: process begins without a specific trigger marker.
- GLAM meaning: internal kickoff (project approved; package created).
- Paradata payload: initiation authority; project ID; scope statement.

Start event (message) [second left].

- BPMN meaning: process starts upon receiving a message.
- GLAM meaning: digitization begins upon intake request or package arrival from a partner.
- Paradata payload: request record; intake channel; requester identity.

Intermediate event (timer) [central].

- BPMN meaning: waits for a specific time/date or cycle; boundary variants can interrupt or not interrupt attached activity.
- GLAM meaning: "waiting states" (cool-down time; scheduled verification; embargo end; periodic fixity checks; waiting for repository validation), including non-interrupting reminders versus interrupting deadlines.
- Paradata payload: schedule; due dates; timeout outcomes.

Intermediate event (message) [second right].

- BPMN meaning: catches or throws a message in the flow; may be source or target for message flow but not both.
- GLAM meaning: asynchronous communications that gate progress (QA request/response; curator approval; partner clarification).
- Paradata payload: message thread; correlation to package; response time.

Intermediate event (escalation) [right].

- BPMN meaning: can be interrupting or non-interrupting; escalation handling without necessarily aborting all threads.
- GLAM meaning: "issue raised" (rights ambiguity; object too fragile; suspected sensitive content) that may trigger parallel governance action while other work continues.
- Paradata payload: severity; escalation path; decision authority.

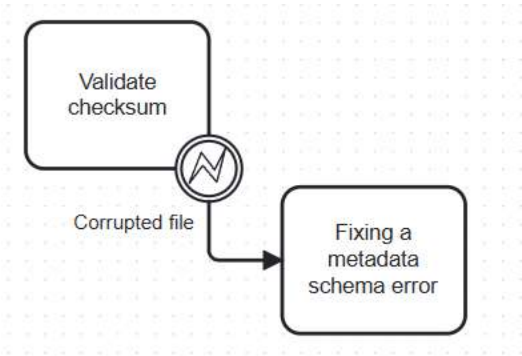


Fig. 7. BPMN intermediate error event symbol indicating an interrupting error used for exception handling.

Intermediate event (error).

- BPMN meaning: boundary catch error interrupts the activity; used for exception handling.
- GLAM meaning: hard failure (checksum mismatch; corrupted file; failed ingest validation; hardware fault) that interrupts the current step and routes to remediation.
- Paradata payload: error code; diagnostics; affected artifacts; fix actions.

Fig. 8. Three BPMN end event symbols showing a message end event, a terminate end event, and a compensation end event.

End event (message).

- BPMN meaning: sends a message at process conclusion.
- GLAM meaning: completion notification to partner/platform (deposit complete; DOI minted; publication live).
- Paradata payload: confirmation artifact; recipients; persistent IDs.

End event (terminate).

- BPMN meaning: immediately ends all activities in the process without compensation/handling.
- GLAM meaning: hard stop (project cancellation; legal takedown requiring immediate halt).
- Paradata payload: authority; reason; scope of termination.

End event (compensation).

- BPMN meaning: triggers compensation for previously completed activities when defined.
- GLAM meaning: structured rollback (withdraw published derivatives; revert access settings; replace incorrect metadata release).
- Paradata payload: compensated steps; reversal evidence; updated versions.

Fig. 9. BPMN boundary event symbols illustrating interrupting and non-interrupting variants attached to an activity.

Boundary event (interrupting vs non-interrupting).

- BPMN meaning: attached to an activity; interrupting cancels the activity; non-interrupting spawns an exception flow while the activity continues.
- GLAM meaning: deadlines, alerts, and issues attached to long-running work (e.g., “rights review overdue” interrupting publication; “QA warning” non-interrupting with parallel annotation).
- Paradata payload: trigger; attachment target; exception flow outcomes.

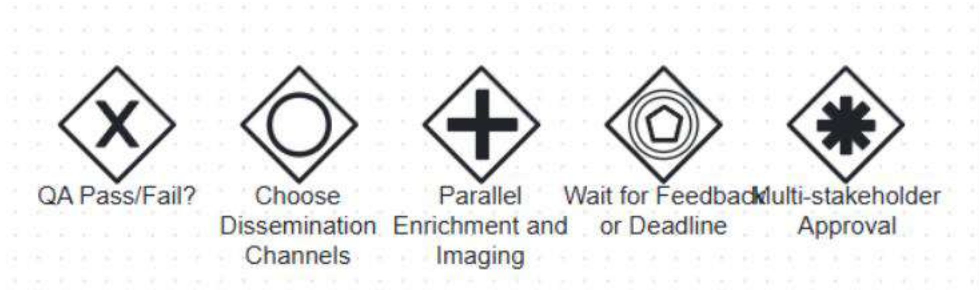


Fig. 10. BPMN gateway symbols showing exclusive (XOR), inclusive (OR), parallel (AND), event-based, and complex gateway types.

Exclusive gateway (XOR).

- BPMN meaning: selects one path based on conditions.
- GLAM meaning: decision points (QA pass/fail; choose capture method; choose dissemination format; license choice).
- Paradata payload: decision rule; evidence; decision-maker role.

Inclusive gateway (OR).

- BPMN meaning: one or more paths may be taken based on conditions.
- GLAM meaning: choose multiple dissemination channels (publish to local portal and Zenodo; produce both preservation and access copies).
- Paradata payload: selection criteria; channel list; consistency checks.

Parallel gateway (AND).

- BPMN meaning: fork/join for parallel execution.
- GLAM meaning: parallelizable work (metadata enrichment in parallel with imaging; derivative generation in parallel with documentation; multi-platform deposit).
- Paradata payload: concurrency constraints; synchronization condition; partial completion handling.

Event-based gateway.

- BPMN meaning: branching based on which event occurs first (message/timer/etc.).
- GLAM meaning: “wait for external response or timeout” (await partner feedback; if none by deadline, escalate or proceed with default).
- Paradata payload: awaited events; timeout policy; chosen outcome.

Complex gateway.

- BPMN meaning: advanced synchronization/branching logic.
- GLAM meaning: exceptional coordination patterns (multi-stakeholder approvals; conditional joins based on partial deliveries).
- Paradata payload: formalized join criteria; justification.

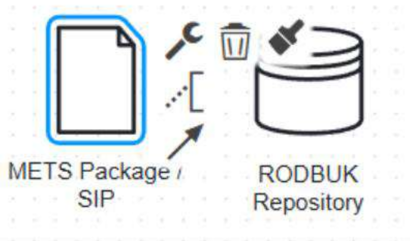


Fig. 11. METS SIP data object and RODBUK repository shown as BPMN data object and data store symbols.

Data object.

- BPMN meaning: represents data produced/required; an "item-aware" construct for things used and created during execution.
- GLAM meaning: concrete digital artifacts at any granularity: single image, OCR output, 3D mesh, metadata record export, QA report, checksum manifest, zipped submission package (SIP/DIP), derivative sets.
- Paradata payload: identifier; format; size; fixity; provenance pointers.

Data store.

- BPMN meaning: persistent storage available to the process.
- GLAM meaning: systems-of-record and repositories (collection management system; preservation storage RODBUK; staging SharePoint/NAS; registry).
- Paradata payload: system name/version; retention policy; access control; preservation guarantees.

Data input and data output.

- BPMN meaning: formal input/output specifications of activities including tasks and services.
- GLAM meaning: declared interface of each step (what files/metadata are required; what products are emitted), which is essential for reproducibility in scientific tiers.
- Paradata payload: schema references; validation constraints; mapping notes.

ItemDefinition and "hooks" for external data structure.

- BPMN meaning: BPMN does not model data structure internally; it formalizes hooks for externally defined data structures and expression languages; XML Schema/XPath are defaults.
- GLAM meaning: attach domain ontologies and standards as the semantic layer for paradata/metadata (e.g., CRMdig for capture events, PREMIS for preservation events, PROV-O for agent/activity provenance) while BPMN provides the procedural spine.
- Paradata payload: ontology identifiers; schema URIs; expression language; mapping version.

9.3.3. Task-Type Guidance for GLAM Digitization

Typical GLAM task-type guidance anchored in BPMN semantics: Repositioning a camera, placing calibration targets, manipulating an object, or packaging for transport should be modelled as Manual Tasks when the work is embodied and not application-mediated, consistent with protocols emphasizing trained handling and controlled physical procedures.

Choosing settings in capture/processing software, editing scans, entering metadata in a UI, approving QA results in a dashboard, and setting rights/access fields should be modeled as User Tasks because they are human actions performed with software assistance and typically managed via task lists or runtime assignment.

Automated runs such as batch conversion, OCR, photogrammetry reconstruction, derivative generation, and metadata normalization should be modelled as Script Tasks when executed as scripts under engine control, or Service Tasks when invoked as external services, and policy-driven determinations (license selection logic, tier assignment, privacy redaction thresholds) should be modelled as Business Rule Tasks.

10. Legal Aspects

10.1. Introduction

Paradata is not a concept that is widely used in legal scholarship or practice, as evidenced by the scarcity of literature on the topic[1]. That, however, does not mean that paradata as such are not used or produced in legal contexts. Lawyers create large amounts of data in their daily work: they gather facts, build databases of legal knowledge that may include annotations, and apply legal rules to facts, thereby producing analytical outputs such as legal assessments or conclusions. At the same time, the processes of data collection and processing in legal practice may themselves be governed by law. Straightforward examples in the EU include the processing of personal data under the GDPR[2], as well as, to a certain extent, the training and use of AI systems covered by the AI Act[3]. Either EU law or national laws may therefore require managing legal obligations in a manner that at least encourages or even mandates tracking of the processes and practices behind data collection and processing.

In this context, paradata may be useful from different perspectives. First, the ability to reconstruct the reasoning or decision-making process after some time can be practically valuable as documentation. Second, paradata may be required to demonstrate compliance with specific legal or regulatory requirements. Below we describe three examples of legal use-cases of paradata in the contexts relevant for the IMPULSE project.

10.2. Use-case 1: paradata and regulatory requirements

Cultural heritage objects may include personal data, i.e. data about natural persons who are identified or identifiable. Furthermore, personal data may also be included in metadata. Since in typical cases, at least some obligations under the GDPR apply to processing of such data[4], under Article 5(2) GDPR the data controller is required to demonstrate compliance with the principles relating to such processing. This *principle of accountability* covers all the other general principles (lawfulness, fairness, transparency,

[1] Among the few contributions see especially Enquist L (2024) *Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes* In: Huvila I, Andersson L, Sköld O (eds) Perspectives on Paradata. Research and Practice of Documenting Process Knowledge, Springer: 233:247

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016, p. 1–88)

[3] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (OJ L, 2024/1689, 12.7.2024)

[4] As discussed in IMPULSE Deliverable D4.1: Legal issues and scopes of activity, pp. 28-29.

purpose limitation, data minimisation, accuracy, storage limitation, integrity and confidentiality) and, by extension, the more specific obligations that originate from them.

Although there is no single correct way of applying the principle of accountability in practice, there are concrete examples in the GDPR. They are accompanied by guidelines that influence the enforcement[5] which include examples of common accountability measures, some of which lead essentially to creating paradata. As such, the measures may include:

- maintaining records of processing activities (Article 30 GDPR), which, in essence, must say what data, how and to what ends are being processed,
- internal procedures and policies to be considered and applied to new data processing operations,
- procedures that manage access, correction and deletion requests, as well as other requests by the data subject.

In this sense, although the GDPR does not require the creation of paradata, paradata emerge as a practice-based consequence, when a controller complies with the principle of accountability.

10.3. Use-case 2: verification of accuracy of metadata

The status of digitizations of cultural heritage objects under copyright law is notoriously difficult to ascertain and controversial[6]. The problem persists despite the efforts to clarify the situation by introducing Article 14 of the 2019 CDSM Directive[7]. According to this provision, when a reproduction is made of a public-domain work of visual art, it should not be subject to copyright or related rights, unless the material resulting from that act of reproduction is original in the sense that it is the author's own intellectual creation. However, some claim that copyright covers such reproductions (e.g. photographs of paintings) and CHIs may adopt practices according to such opinions.

At the same time, information about the status of a given representation under intellectual property law should and is often included in the metadata, either in the form of a licensing statement, by reference to a non-binding standard such as Rights Statement[8], or otherwise. What is missing is how a given institution *reached the conclusion displayed in the metadata*. There is no information about the assessment process, the

[5] See in particular Opinion 3/2010 on the principle of accountability (WP173) issued by the Article 29 Data Protection Working Party (the WP was replaced by the European Data Protection Board, but the opinion remains persuasive authority under the GDPR).
[6] As discussed in IMPULSE Deliverable D4.1: Legal issues and scopes of activity, pp. 18-19.
[7] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (OJ L 130, 17.5.2019, p. 92–125).
[8] https://rightsstatements.org

assumptions made, or the legal reasoning that led to the stated copyright status. Let us consider two examples from the collections accessible via Europeana.

Fig. 12. „Liegendes Pferd" by Franz Marc (1880-1918)[9]

The author of *Liegendes Pferd* died in 1918 which means that copyright has lapsed decades ago. However, the heritage institution claims that the reproduction is under copyright and a closer inspection of the metadata (© *Deutsche Fotothek / Margarete Fimmen-Lindig*) suggests that copyright is claimed not to the digital copy as such, but to a photograph of the original painting by Franz Marc taken by M. Fimmen-Lindig, which was later digitised. Even this brief information allows a potential user to verify the accuracy of this interpretation and, if they are willing to risk a dispute with the institution, to contest it.

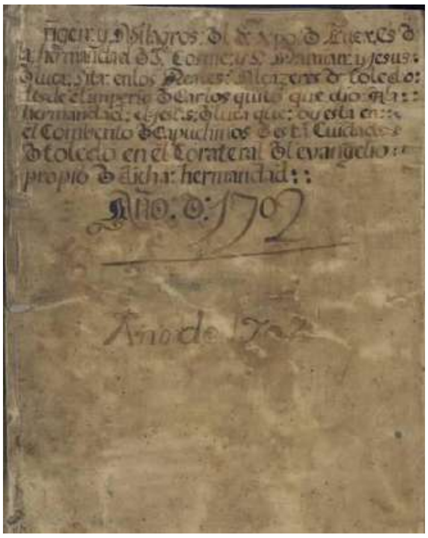


Fig. 13. "De inventione, revelatione ac translatione sacratissimi vultus [de Luca]" (15th century)[10]

[9] From the collection of *Deutsche Fototek*: https://www.deutschefotothek.de/documents/obj/72034552/df_hauptkatalog_0127058

[10] From the collection of Biblioteca Virtual del Patrimonio Bibliografico (https://bvpb.mcu.es/es/consulta/registro.do?id=397681)

Again, the original object is not protected by copyright, since a 15th-century manuscript was made before copyright was introduced to any legal system in the world. A closer inspection of the metadata reveals the institution's claim that the digital copy is available under the Creative Commons BY 4.0 license. This license, although very permissive, assumes nonetheless that there is a work protected by copyright that requires licensing in the first place. Such an assertion is, however, extremely contestable, as the reproduction appears to be a 2D-scan that is not original in the sense attached to this term by EU copyright law, i.e. does not reflect the personality of its author as an expression of their free and creative choices, regardless of the skill and labour that may have been necessary to make it. [11]Conceivably, the institution may have had something else in mind, such as a possibility that the entire database that includes this particular reproduction is protected. However, without a more detailed description it is impossible to verify it without contacting the institution first. Furthermore, it is uncertain whether the conclusion described above as contestable was reached by a legal professional, or after any consultation with a person trained in intellectual property law, or by a layperson who, even in good faith, may have misapplied the law[12]. This is where paradata, providing more detailed information about the processes of preparing legal metadata, would be particularly useful.

10.4. Use-case 3: using paradata to create legal metadata

Finally, paradata support legal analysis as well as making informed decisions about metadata. This appears particularly useful in the context of copyright protection of digitization. As recently explained by the Court of Justice of the European Union, not every choice made by a person aspiring to be an author under copyright law is relevant. "Choices dictated by various constraints, particularly technical, which bound that author during the creation of that subject matter are not free and creative, and nor are choices which, while free, do not bear the imprint of the author's personality by giving that subject matter a unique appearance[13]". Information about such constraints is not self-evident when one considers the digital copy by itself, but paradata may provide the missing information.

For example, the sequence used by Heritage Malta mentions not only the details of the digitization process, but also the rationale of decision-making (e.g. why a certain angle or lighting was chosen), as well as interpretive decisions (e.g. reconstruction of missing parts, 3D modelling assumptions); which may, depending on their exact nature, be relevant in the analysis of whether a given digitization is original according to copyright law, and thus improve the quality of legal metadata.

[11] As discussed in IMPULSE Deliverable D4.1: Legal issues and scopes of activity, pp. 18.
[12] Such a situation may be caused by institutional or budgetary constraints, applying legacy workflows or overreliance on templates.
[13] Cases C-580/23 and C-795/23 Mio/konektra, ECLI:EU:C:2025:941.

11. Recommendations- Paradata Decisions

11.1 A Multi-Dimensional, Three-Tier, Five-Phase Approach

11.1.1 Three Tiers: from the Minimum Viable to the Digital Twin

The cross- examination of paradata workflows originating from different dataset types and featuring different institutional priorities and decisions led us to the development of three common tiers that describe adequately the decision processes for the capture of paradata.

Tier 1: Minimum Viable:

- This tier establishes the essential administrative and technical foundation required to make digital assets accessible for general public use and visualization. It focuses on capturing core project details such as institutional information, object type and dimensions, project identifiers, and digitization purpose (alongside basic metadata standards, language settings, and acquisition methods). Documentation at this level ensures that the digital object can be stored, published, and shared with appropriate rights statements and links to its physical counterpart, providing a functional baseline for outreach and engagement.

Tier 2: Scientific

- The second tier introduces enhanced documentation to support scholarly research and reproducibility. It incorporates quality benchmarks, controlled vocabularies, and provenance details, as well as technical specifications for capture and processing. This includes recording calibration tools, quality assurance standards, and enrichment processes, along with uncertainty levels and preservation formats. By aligning with recognized frameworks and persistent identifiers, this tier ensures that digital assets meet rigorous standards for accuracy, interoperability, and long-term reuse within academic contexts.

Tier 3: Memory Twin

- The highest tier addresses the full complexity of digital heritage representation, emphasizing interpretative transparency and advanced technical detail. It documents material and geometric challenges, environmental conditions during capture, and specialized setups such as lighting and color calibration. This level also tracks manual interventions, enrichment methodologies, and any AI models applied, alongside reduction or redaction logs for public dissemination. By providing a comprehensive paradata report, Tier 3 creates a “memory twin” of the object, enabling future researchers to understand every decision and constraint that shaped its digital form.

Within these 3 tiers, the specific approach to paradata creation and documentation of each institution will be analyzed, to identify any common patterns that can potentially serve as a guiding scheme for a wider pool of cultural institutions dealing with the emerging field of paradata.

11.1.2 Five Phases: from Planning to Publication

A comparative, cross- institutional analysis of the paradata process workflow of each IMPULSE institution led to the “codification” of the process in 5 phases.

Phase 1: Planning & Complexity Assessment

This phase lays the groundwork for the digitization project by defining its scope, objectives, and complexity. The purpose is to capture the digital representation of the object with accuracy and reproducibility. At the most basic level (minimum viable), it involves recording administrative details such as the institution name, object type, size, and project identifiers, along with the purpose of digitization. As the tiers progress, additional considerations are introduced, including quality targets, intended output formats, and compliance with frameworks like the Europeana Publishing Framework (scientific). At the highest tier (Memory twin), complexity assessments become critical, addressing material, geometric, and environmental challenges to ensure that technical and logistical constraints are fully understood before data capture begins.

Phase 2: Metadata Strategy

Metadata strategy ensures that the digital assets are discoverable, interoperable, and trustworthy. In the first tier, this phase specifies the metadata standard (LIDO, MARC21, Dublin Core etc.) and language used for description. Tier two adds layers of version control by cross validating the dates of original and updated metadata, controlled vocabularies (“Getty AAT”, geonames etc.) and thesauri. It also adds the aspects of metadata source (origin on the information of metadata), creator role and the metrics of technical details. At the Memory Twin level, metadata aggregation standards and multilingual support are incorporated to facilitate interoperability across platforms and institutions.

Phase 3: Data Acquisition

Data acquisition focuses on capturing the digital representation of cultural objects with precision and reproducibility. At the minimum viable level, it documents the acquisition method (such as 2D, Photogrammetry or Structured Light), hardware equipment used, and the operator responsible for the process. As complexity increases in the scientific tier, the phase includes detailed quality metrics, calibration tools, quality assessment and adherence to recognized guidelines such as FADGI and Metamorfoze. The memory twin tier introduces lighting setups describing how the object was lit, color calibration, environmental conditions and raw dataset size, ensuring that every factor influencing data quality is recorded.

Phase 4: Processing & Enrichment

This phase transforms raw data into enriched, usable assets while maintaining transparency in the methods applied. Initially at the minimum viable tier it involves noting the processing software used and any optimization steps for user engagement, as such documentation will be particularly useful for the IMPULSE XR requirements. In the scientific tier, processing parameters, i.e. key settings such as “High Quality Alignment’ are documented to ensure reproducibility. “Enrichment type is also captured, referring to the type of post processing. Finally, the uncertainty level is also documented. At the Memory Twin level, the process includes manual intervention logs, enrichment methodologies, and the identification of AI models used.

Phase 5: Ingest & Publication

The final phase ensures long-term preservation and dissemination of digital assets. At the minimum viable tier, it records dataset type, rights statements, and links to original sources. Scientific tier introduces archival and dissemination formats, persistent identifiers, and compliance with standards such as OAIS and PREMIS. At the Memory Twin tier, transparency is enhanced through reduction or redaction logs and links to full paradata reports compliant with the Eureka 3D model. This structured approach guarantees that digital assets remain accessible, sustainable, and legally compliant, supporting both institutional stewardship and public engagement.

11.2. Proposed Paradata Structure

11.2.1. Tier 1: Minimum Viable

Goal: Provides the basic administrative and technical baseline necessary for public use and visualization.

Phase	Field Name	Description / Instruction
1. Planning & Complexity Assessment	Institution	Name/Address/Contact
1. Planning & Complexity Assessment	Object type	object description (e.g. "sculpture" "early print" "carrier format")
1. Planning & Complexity Assessment	Object size(s)	Object size (e.g. "210x297 mm", "1.5x2.3x0.75 m")
1. Planning & Complexity Assessment	Project ID	Unique internal identifier
1. Planning & Complexity Assessment	Project Scope	Scope of the project, including funding body, ...
1. Planning & Complexity Assessment	Digitization Purpose	Primary goal (e.g. "General Outreach" "Conservation" "VR Experience" "Research").
2. Metadata Strategy	Metadata Standard	Standard used for description in local system (e.g. "LIDO v1.0" "Dublin Core" "MARC21").
2. Metadata Strategy	Language of Description	Primary language(s) of the metadata (e.g. "EN" "NL" "FR").
3. Data Acquisition	Acquisition Method	Technique used (e.g. "2D" "Photogrammetry" "Structured Light").
3. Data Acquisition	Equipment Hardware	Specific model (e.g. "Sony A7R IV + 50mm Lens" "Leica BLK360").
3. Data Acquisition	Operator Name	Person responsible for the capture event.
4. Processing & Enrichment	Processing Software	Software & Version used (e.g. "RealityCapture 1.2" "Metashape Pro 1.8" "ABBYY Finereader" "ImageJ").
4. Processing & Enrichment	Optimization for user engagement (reduction)	Logic for reduction (e.g. "Decimated to 50k polys for compatibility").
5. Ingest & Publication	Dataset Type	Preservation Master (Full Data) or "Consultation Copy" (Reduced/Public).
5. Ingest & Publication	Rights Statement (Data)	License for the digital data (e.g. "CC-BY 4.0" "Public Domain Mark").

5. Ingest & Publication	Link to Original Source	Identifier/URL of the physical object or archival record (e.g. "Inventory No. 1234").

11.2.2. Tier 2: Scientific

Goal: Adds critical metrics for accuracy and standard alignment, making the asset suitable for scholarly research and basic reproduction.

Phase	Field Name	Description / Instruction
1. Planning & Complexity Assessment	Target Quality	High Quality Reuse, Desired Europeana Tier (e.g. Tier 4)
1. Planning & Complexity Assessment	Intended Output Format	The target delivery format (e.g. "High-Poly OBJ","Web-Ready glTF", "JPG", "MP4")
2. Metadata Strategy	Metadata versioning	Date of original metadata, date of updated metadata
2. Metadata Strategy	Thesauri / Vocabularies	Controlled vocabularies used (e.g. "Getty AAT" "geonames.org" "Local Authority File").
2. Metadata Strategy	Source of Metadata	Origin of the information (e.g. "Physical Card Catalog" "Legacy Database" "Digital Born").
2. Metadata Strategy	Metadata Creator Role	Role of the agent creating metadata (e.g. "Curator" "Archivist" "AI Agent").
3. Data Acquisition	Capture Quality	Metric of detail (e.g. "600 dpi" "GSD: 0.1mm/pixel" "Point Density: 5mm @ 10m" "4K", "Dolby Surround").
3. Data Acquisition	Quality Assessment	Standard or guideline used for quality assurance (e.g. "Metamorfoze" "FADGI")
3. Data Acquisition	Metric Calibration	Tool used for scale (e.g. "Invar Scale Bars" "Coded Targets").
4. Processing & Enrichment	Processing Parameters	Key settings (e.g. "High Quality Alignment" "Mild Depth Filtering" "PCA").
4. Processing & Enrichment	Enrichment Type	type of postprocessing ("OCR" "false color visualisation")
4. Processing & Enrichment	Uncertainty Level	Confidence in the reconstruction (Scale 1-5: 1=Hypothetical 5=Measured).

5. Ingest & Publication	Preservation Format (SIP)	Format of the archival file (e.g. "E57" "OBJ + High-Res TIFF").
5. Ingest & Publication	Dissemination Format (DIP)	Format for web/XR viewing (e.g. "glTF" "USDZ").
5. Ingest & Publication	Persistent Identifier	DOI ARK or Handle for the dataset.

11.2.3. Tier 3: Memory Twin

Goal: Capture the full complexity.

Phase	Field Name	Description / Instruction
1. Planning & Complexity Assessment	Object Material Complexity	Assessment of surface challenges (e.g. "High Reflectivity" "Transparency" "Uniform Texture").
1. Planning & Complexity Assessment	Object Geometric Complexity	Assessment of structural challenges (e.g. "Deep Cavities" "Occluded Surfaces" "Moving Parts").
1. Planning & Complexity Assessment	Site/Environment Complexity	Constraints at capture site (e.g. "No Tripod Allowed" "Low Light" "Public Access").
2. Metadata Strategy	Metadata Standard 2	Standard used for description after aggregation (e.g. "Dublin Core", "EDM")
3. Data Acquisition	Lighting Setup	How the object was lit (e.g. "Cross-Polarized Flash" "Natural Diffuse" "Ring Light").
3. Data Acquisition	Color Calibration	Tool used for color reference (e.g. "X-Rite ColorChecker SG" "SpyderCheckr").
3. Data Acquisition	Environmental Conditions	Relevant factors at time of capture (e.g. "Cloudy" "Wind Speed 10km/h" "Indoor Controlled").
3. Data Acquisition	Dataset Raw Size	Count of source files (e.g. "450 RAW Images" "12 Scan Positions").
4. Processing & Enrichment	Manual intervention Log	Intervention type (e.g. "OCR correction" "Metadata cleanup" "Noise Reduction")
4. Processing & Enrichment	Enrichment Methodology	Manual, Automated, Hybrid (Human-in-the-Loop)

4. Processing & Enrichment	AI Model Used (if any)	Name/Version of algorithm (e.g. "BERT-base-multilingual" "NeRF").
5. Ingest & Publication	Reduction/Redaction Log	Description of what was removed for the public copy (e.g. "Admin data removed" "Faces blurred").
5. Ingest & Publication	Paradata Report Link	URL to the full paradata log (if external).

11.3 Mapping the BPMN model to the Proposed Paradata Approach based on the five phases

11.3.1 Phase 1: Planning & Complexity Assessment

Each one of the sections below describes one of the five phases of the digitisation workflow presented above, mapped into the BPMN model. Each sub-section includes a BPMN diagram, a diagram description, as well as an example of the BPMN notation in XML.

Diagram

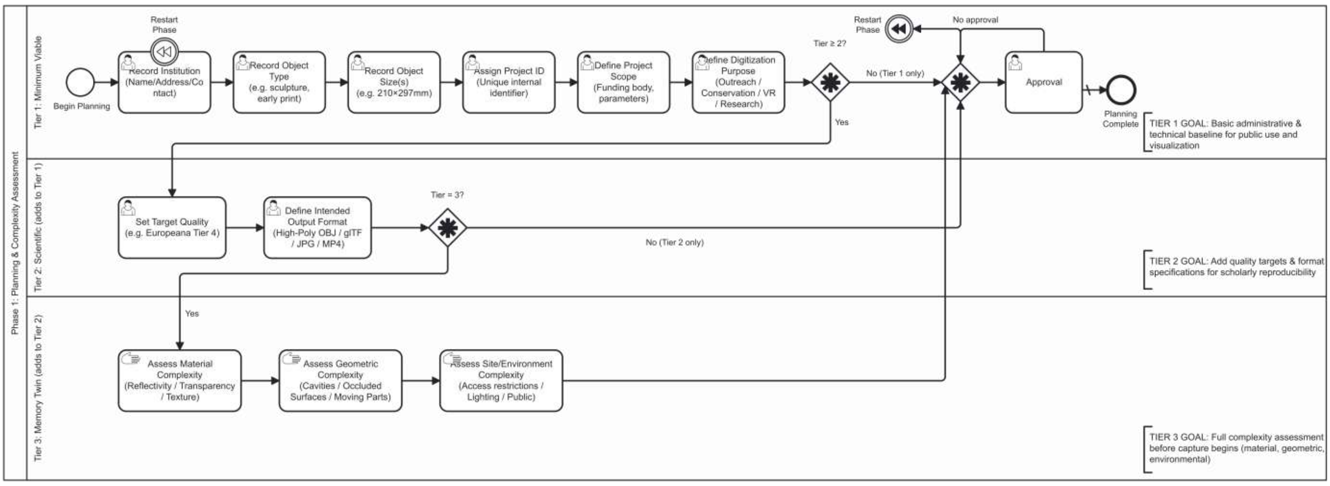


Figure 14. BPMN diagram corresponding to "Phase 1- Planning and Complexity Assessment".

Diagram Description

The Phase 1 BPMN diagram illustrates the planning and complexity assessment workflow using a three-tier swim lane structure (Tier 1: Minimum Viable, Tier 2: Scientific, Tier 3: Memory Twin). The process employs conditional execution via Complex Gateways to route projects through appropriate tasks based on their selected tier. All projects complete Tier 1 tasks (institution details, object type, size, project ID, scope, and purpose). Projects selecting Tier 2 or higher proceed to define target quality and output formats. Additionally, Tier 3 projects undergo a comprehensive assessment of complexity across three dimensions: material (reflectivity, transparency, texture), geometric (cavities, occluded surfaces, moving parts), and environmental (site access, lighting, safety protocols). The workflow uses User Tasks for decision-making activities and Boundary Events to enable restarts, ensuring projects can return to earlier stages when fundamental issues are identified. This tier-based conditional flow ensures that documentation depth scales appropriately with project ambition and research requirements.

Example Rationale: Medieval Castle Ruins (Tier 3)

We selected the Institute of Archaeology's medieval castle ruins project (PROJ-2025-007) as our focal example because it represents the most complete and challenging scenario an institution may encounter in Phase 1 planning. This project required three submission attempts before approval, demonstrating the iterative nature of Tier 3 planning and the depth of documentation required.

Why focus on Tier 3 complex cases? Tier 3 "Memory Twin" projects represent the highest standard of digital heritage documentation, requiring complete provenance, risk assessment, and reproducibility planning. By examining a complex Tier 3 case that ultimately achieves approval, we provide evaluators with a comprehensive template that encompasses all possible planning challenges: quantified complexity assessment (40% moss coverage, 70% accessible surfaces, 30% inaccessible), regulatory compliance (drone permits from Polish Civil Aviation Authority, structural engineer sign-offs for scaffolding access), safety protocols (PPE requirements, hospital proximity), environmental contingencies (weather windows with backup dates), and budget realism (€85,000 with 14% contingency).

An approved Tier 3 case serves as the "complete playbook" - institutions facing simpler Tier 1 or Tier 2 projects can reference the comprehensive Tier 3 standard and identify which elements apply to their context. The castle ruins example shows that even ambitious, complex projects succeed when planning is sufficiently detailed, risks are systematically addressed, and contingencies are built in from the start. The three-attempt progression (vague → improved → exemplary) also demonstrates the learning curve institutions experience when first attempting Tier 3 documentation, providing realistic expectations for project timelines and the value of detailed reviewer feedback.

Paradata example based on the Planning and Complexity phase, according to BPMN notation

<Tier1Data>
<Institution>Institute of Archaeology and Ethnology, Polish Academy of Sciences, ul. Rubież 46, 61-612 Poznań. Contact: Prof. Marek Kamiński, kaminski@iae.edu.pl, +48 61 829 5600**</Institution>**

```xml
            <ObjectType>Archaeological site - medieval castle ruins (13th
            century). Masonry construction: limestone/sandstone, partially
            collapsed defensive towers, remnant curtain walls.</ObjectType>
            <ObjectSize>Site footprint: 45m × 30m. Wall heights: 4-12m (variable
            preservation). Tower remnants: max 12m height. Interior courtyard:
            20m × 15m.</ObjectSize>
            <ProjectID>IAE-2025-CASTLE-007-TIER3</ProjectID>
            <ProjectScope>Complete archaeological site documentation for dual
            output: (1) Public VR educational experience (museum installation),
            (2) Research publication on medieval fortification architecture. EU
            Horizon Europe grant: €85,000 (breakdown: Phase 1-2: €8K, Phase 3
            acquisition: €45K, Phase 4-5: €20K, contingency: €12K). Timeline: 6
            months (March-August 2025). Team: 3 archaeologists, 2
            photogrammetrists, 1 VR developer.</ProjectScope>
            <DigitizationPurpose>Dual purpose: (1) Public education - VR
            experience for museum visitors to explore reconstructed castle, (2)
            Research - Publication in "Medieval Archaeology" journal analyzing
            defensive architecture evolution</DigitizationPurpose>
      </Tier1Data>
      <Tier2Data>
            <TargetQuality>Europeana Tier 4 (research-grade). Target accuracy:
            ±2mm for accessible surfaces, ±5cm for drone-captured elevated
            features. Ground Sample Distance (GSD): 1mm for close-range, 5mm for
            aerial.</TargetQuality>
            <IntendedOutputFormat>Multi-resolution deliverables: (1) Archival:
            Dense point cloud (200M+ points, E57), High-poly mesh (20M triangles,
            PLY/OBJ, 16K textures). (2) Research: Medium-poly (5M triangles) with
            annotations. (3) VR: Optimized (2M triangles, FBX for Unity, 4K
            textures, LOD system).</IntendedOutputFormat>
            </Tier2Data>
      <Tier3Data>
            <MaterialComplexity>
                  <Assessment>Challenging (4 on 5-point scale)</Assessment>
                  <StoneType>Mixed: Limestone (load-bearing walls) and sandstone
                  (decorative elements). Variable weathering: 30% well-preserved,
                  40% moderate erosion, 30% heavily weathered.</StoneType>
                  <BiologicalGrowth>~40% surface covered with moss/lichen
                  (primarily north-facing walls). Professional conservator
                  engaged for non-destructive removal (February 2025, €3,500
                  allocated).</BiologicalGrowth>
                  <OpticalChallenges>High reflectivity variation (wet stone after
                  rain vs dry stone - factor 3x difference). Dark shadows in deep
                  window embrasures and interior spaces.</OpticalChallenges>
                  <ColorCalibration>X-Rite ColorChecker Passport will be placed
                  in each capture session. Calibration targets at 5m intervals
                  around site perimeter for consistent color across multi-day
                  capture.</ColorCalibration>
            </MaterialComplexity>
            <GeometricComplexity>
                  <Assessment>Very Complex (5 on 5-point scale)</Assessment>
                  <AccessibilityBreakdown>Accessible: 70% (exterior walls,
                  courtyard, safe ground-level features). Limited access: 20%
                  (requires scaffolding - interior vaults, upper wall sections).
                  Inaccessible: 10% (collapsed tower interiors - structurally
                  unsafe per engineer assessment).</AccessibilityBreakdown>
                  <StructuralFeatures>Deep window embrasures (1.2m depth), spiral
                  staircase remnants (22 steps visible), Gothic vaulted ceiling
                  fragments (ribs preserved), murder holes, arrow
                  slits.</StructuralFeatures>
```

```xml
            <OcclusionStrategy>Multi-modal capture: (1) Terrestrial
            photogrammetry (300+ positions), (2) Drone (DJI Mavic 3 - CAA
            permission obtained, flight plan approved, see doc #IAE-2025-
            DRONE-PERMIT), (3) Interior scaffolding (contractor: SafeScaff
            Ltd, structural engineer sign-off: Dr. Eng. Piotr Zieliński,
            dated 2024-12-15).</OcclusionStrategy>
        </GeometricComplexity>
        <EnvironmentComplexity>
            <Assessment>Moderate to High (4 on 5-point scale)</Assessment>
            <SiteStatus>Public archaeological park (15,000 visitors/year).
            Managed by National Heritage Board. Capture permit: NHB-2025-
            RESEARCH-089 (valid March-May 2025).</SiteStatus>
            <AccessRestrictions>Site closed to public during capture (March
            15 - April 30). Security: Park rangers on-site. Equipment
            staging area: parking lot 50m from site.</AccessRestrictions>
            <VegetationManagement>Invasive vegetation removal: February
            2025 (contractor: GreenCare Conservation, €2,800). Mature oak
            trees around perimeter cannot be removed (protected status) -
            will cast morning shadows on east wall (capture plan adjusted
            for afternoon light).</VegetationManagement>
            <WeatherStrategy>Optimal capture window: April 1-30, 2025.
            Target: Overcast days (even lighting, no harsh shadows).
            Weather monitoring: Daily forecasts, 7-day rolling plan.
            Contingency: If April unfavorable, extend to May 15 (before
            peak tourism). Team on standby for optimal
            conditions.</WeatherStrategy>
            <SafetyAssessment>Structural stability assessment completed by
            Dr. Eng. Piotr Zieliński (2024-12-15). Safe access zones
            marked. Scaffolding design approved for interior vault access
            (load limit: 200kg, max 2 persons). Personal protective
            equipment (PPE): Hard hats, safety harnesses for elevated work.
            First aid: Team member certified, hospital 12km
            away.</SafetyAssessment>
            <RegulatoryCompliance>Drone flight permit: Polish Civil
            Aviation Authority (CAA) approval #DRONE-2025-IAE-007, valid
            March-May. Site within 8km of regional airport - flights
            coordinated with air traffic control. Insurance: Professional
            liability €2M, equipment €50K.</RegulatoryCompliance>
        </EnvironmentComplexity>
    </Tier3Data>
    <Approval>
        <ApprovedBy>Project Director - Prof. Marek Kamiński</ApprovedBy>
        <ApprovalDate>2025-01-14</ApprovalDate>
        <Comments>APPROVED: Tier 3 planning now comprehensive and realistic.
        All complexity dimensions thoroughly assessed with quantified
        metrics. Safety documentation complete (structural engineer sign-off
        critical). Regulatory compliance addressed (drone permits, site
        access). Budget breakdown realistic with contingency. Weather
        strategy includes contingency plan. Color calibration strategy
        appropriate. This level of planning detail significantly reduces
        Phase 3 capture risks. Exemplary Tier 3 planning documentation. Ready
        for Phase 2: Metadata Strategy.</Comments>
    </Approval>
```

11.3.2 Phase 2. Metadata Strategy

Diagram

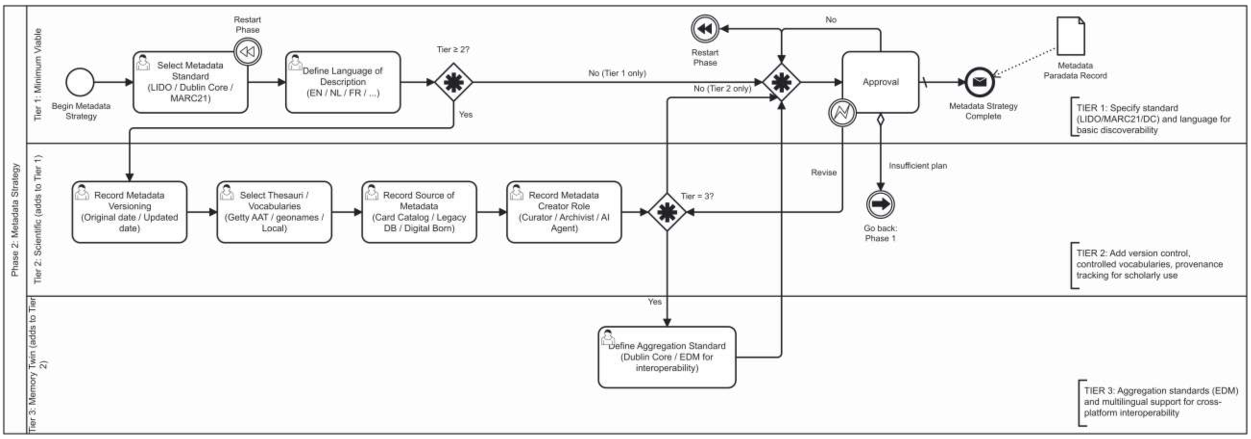


Figure 15. BPMN diagram corresponding to “Phase 2: Metadata Strategy”.

Diagram Description

The Phase 2 BPMN diagram illustrates the metadata strategy workflow using a three-tier swim lane structure (Tier 1: Minimum Viable, Tier 2: Scientific, Tier 3: Memory Twin). The process employs conditional execution via Complex Gateways to route projects through appropriate tasks based on their selected tier. All projects complete Tier 1 tasks: Select Metadata Standard (LIDO / Dublin Core / MARC21) and Define Language of Description (EN / NL / FR / ...). Projects selecting Tier 2 or higher proceed to Record Metadata Versioning (Original date / Updated date), Select Thesauri / Vocabularies (Getty AAT / geonames / Local), Record Source of Metadata (Card Catalog / Legacy DB / Digital Born), and Record Metadata Creator Role (Curator / Archivist / AI Agent). Tier 3 projects additionally complete Define Aggregation Standard (Dublin Core / EDM for interoperability), ensuring comprehensive metadata strategy for cross-repository dissemination. The workflow uses User Tasks for decision-making activities and Complex Gateways (Tier ≥ 2?, Tier = 3?) for conditional routing, ensuring that metadata planning depth scales appropriately with project requirements and long-term interoperability goals.

Example Rationale: Medieval Castle Ruins - LIDO Strategy (Tier 3)

We selected the Institute of Archaeology's medieval castle ruins metadata strategy (PROJ-2025-007) as our focal example because it represents the most complete and challenging metadata scenario an institution may encounter in Phase 2. This project required three submission attempts before approval, demonstrating the depth of documentation required for Tier 3 metadata strategy and the iterative refinement of standards selection, multilingual workflows, and aggregation planning.

Why focus on Tier 3 complex cases? Tier 3 "Memory Twin" projects require comprehensive metadata strategy that ensures long-term interoperability, reproducibility, and multi-repository dissemination. By examining a complex Tier 3 case that ultimately achieves approval, we provide evaluators with a comprehensive template that encompasses all possible metadata challenges: primary standard selection with clear relationships (LIDO v1.1 as primary, full descriptive metadata, with Dublin Core Simple derived automatically for basic web discovery), multilingual strategy with documented translation workflows (Polish primary/authoritative, English complete translation, German abstract/summary with sequential process and quality control), controlled vocabularies with persistent URIs (Getty AAT with specific URIs like 300014126 for masonry, GeoNames ID 3088825, PeriodO p0m63nj), complete LIDO→EDM crosswalk documentation with field-level mapping, comprehensive metadata update policy (minor corrections/factual errors/major revisions with versioning), metadata rights strategy (CC0 for metadata, CC BY 4.0 for descriptive text), and ORCID identifiers for all creators ensuring persistent attribution.

An approved Tier 3 metadata strategy serves as the "complete playbook" - institutions facing simpler Tier 1 (basic Dublin Core) or Tier 2 (LIDO with vocabularies) projects can reference the comprehensive Tier 3 standard and identify which elements apply to their context. The castle ruins example shows that even ambitious multilingual, multi-repository metadata strategies succeed when standards relationships are clearly defined, translation workflows are documented with quality control, aggregation crosswalks are complete at field level, and update policies address all contingencies. The three-attempt progression (vague standards relationships → improved but missing workflows → exemplary with full documentation) demonstrates the learning curve institutions experience when first attempting Tier 3 metadata strategy, providing realistic expectations for the iterative nature of comprehensive metadata planning.

Example

```
<Tier1Data>

    <MetadataStandard>LIDO v1.1 (primary, full descriptive metadata) with Dublin Core Simple (derived automatically for basic web discovery)</MetadataStandard>

    <LanguageOfDescription>Polish (pl) primary/authoritative, English (en) complete translation, German (de) abstract/summary only (200 words max)</LanguageOfDescription>

</Tier1Data>

<Tier2Data>

    <MetadataVersioning>

        <InitialCreation>2025-01-14(baseline metadata by curator)</InitialCreation>

        <TranslationComplete>2025-01-18 (English translation added)</TranslationComplete>

        <GermanSummary>2025-01-19 (German abstract added)</GermanSummary>

        <TechnicalReview>2025-01-21 (LIDO validation, controlled vocabulary check)</TechnicalReview>

        <CurrentVersion>v2.0 (2025-01-23)</CurrentVersion>

    </MetadataVersioning>

    <ThesauriVocabularies>

        <MaterialTerms>Getty AAT (masonry: 300014126, limestone: 300011286, sandstone: 300011327)</MaterialTerms>

        <GeographicTerms>GeoNames ID: 3088825 (city), TGN: 7012149 (historical region)</GeographicTerms>

        <ArchitecturalTerms>Local vocabulary: "Polish Medieval Fortifications Terminology" v2023.1. Persistent URI: https://iae.edu.pl/vocab/fortifications/ (maintained by Institute)</ArchitecturalTerms>

        <PeriodTerms>PeriodO: p0m63nj (13th century Poland)</PeriodTerms>

        <Note>All terms include persistent identifiers (URIs) for machine-actionable interoperability</Note>

    </ThesauriVocabularies>
```

<MetadataSource>Integrated from: (1) New field documentation (2025), (2) Archaeological excavation reports (1985: Nowak et al., 1998: Kowalski), (3) Historical primary source: Jan Długosz Chronicle (15th c.), (4) Conservation assessment (2010: Heritage Institute Report #2010-45)**</MetadataSource>**

<MetadataCreatorRole>

<ContentAuthority>Prof. Marek Kamiński (ORCID: 0000-0001-2345-6789) - archaeological content, historical interpretation**</ContentAuthority>**

<MetadataEncoding>Dr. Anna Lewandowska (ORCID: 0000-0002-3456-7890) - LIDO encoding, controlled vocabulary application**</MetadataEncoding>**

<Translation>Dr. Thomas Weber (certified translator EN-DE-PL) - English full translation, German summary**</Translation>**

<QualityControl>Dr. Maria Zawadzka (metadata quality review, LIDO validation)**</QualityControl>**

</MetadataCreatorRole>

</Tier2Data>

<Tier3Data>

<AggregationStandard>

<Standard>EDM (Europeana Data Model) v5.2.8 for aggregation to Europeana Cultural Heritage portal**</Standard>**

<TargetAggregator>Europeana via Polish National Aggregator (NIMOZ)**</TargetAggregator>**

<LIDOtoEDMCrosswalk>

<DescriptiveMetadata>LIDO lido:descriptiveMetadata → EDM ore:Proxy (CHO - Cultural Heritage Object)**</DescriptiveMetadata>**

<TechnicalMetadata>LIDO lido:resourceSet → EDM edm:WebResource (digital representation)**</TechnicalMetadata>**

<RightsInfo>LIDO lido:rightsResource → EDM edm:rights**</RightsInfo>**

<Provenance>LIDO lido:eventSet (acquisition) → EDM edm:ProvidedCHO/dc:provenance**</Provenance>**

<GeospatialData>LIDO lido:geonames → EDM edm:Place (with coordinates)**</GeospatialData>**

<DocumentationURL>Full crosswalk table: https://iae.edu.pl/metadata/lido-edm-crosswalk-v2.0.pdf**</DocumentationURL>**

```xml
        </LIDOtoEDMCrosswalk>

    </AggregationStandard>

    <TranslationWorkflow>

        <Process>Sequential: (1) Polish authoritative metadata created by curator, (2) LIDO encoding by specialist, (3) Translation to English by certified translator, (4) German summary extracted from English by translator, (5) Curator review of translations for technical accuracy</Process>

        <QualityControl>All translations reviewed by content authority (Prof. Kamiński) to ensure archaeological terminology accuracy</QualityControl>

        <Timeline>Polish: Day 1, LIDO encoding: Day 2-3, English translation: Day 4-5, German summary: Day 6, Review: Day 7</Timeline>

    </TranslationWorkflow>

    <MetadataUpdatePolicy>

        <MinorCorrections>Typos, formatting: Update in place, increment patch version (v2.0.1), log in change history</MinorCorrections>

        <FactualCorrections>Errors in dates, attributions: Update with new minor version (v2.1), add correction note with date and reason, notify aggregators</FactualCorrections>

        <MajorRevisions>Substantial reinterpretation: New major version (v3.0), previous version archived with persistent URI, DOI updated, formal erratum published</MajorRevisions>

        <ChangeLog>All changes logged at: https://iae.edu.pl/metadata/castle-007/changelog with timestamps, editor, reason</ChangeLog>

    </MetadataUpdatePolicy>

    <MetadataRightsStrategy>

        <MetadataLicense>CC0 1.0 Universal (Public Domain Dedication) - metadata freely reusable</MetadataLicense>

        <DescriptiveText>Copyright Institute of Archaeology (authorship acknowledged) - CC BY 4.0</DescriptiveText>

        <DigitalObjects>CC BY-NC-SA 4.0 (non-commercial, share-alike for 3D models/images)</DigitalObjects>

        <RightsStatement>Rights statements comply with RightsStatements.org and Europeana Publishing Framework</RightsStatement>

    </MetadataRightsStrategy>
```

</Tier3Data>

<Approval>

<ApprovedBy>Metadata Director - Dr. Maria Zawadzka**</ApprovedBy>**

<ApprovalDate>2025-01-24**</ApprovalDate>**

<Comments>APPROVED: Tier 3 metadata strategy now exemplary. All issues resolved: (1) Multilingual workflow documented with timeline and QC, (2) Local vocabulary has persistent URI, (3) Complete LIDO→EDM crosswalk with documentation link, (4) Comprehensive update policy (minor/major/factual corrections), (5) Rights strategy for metadata, text, and digital objects. Use of ORCIDs for creators excellent practice. This metadata strategy ensures long-term interoperability and supports responsible curation. Ready for Phase 3: Data Acquisition.**</Comments>**

</Approval>

11.3.3 Phase 3: Data Acquisition

Diagram

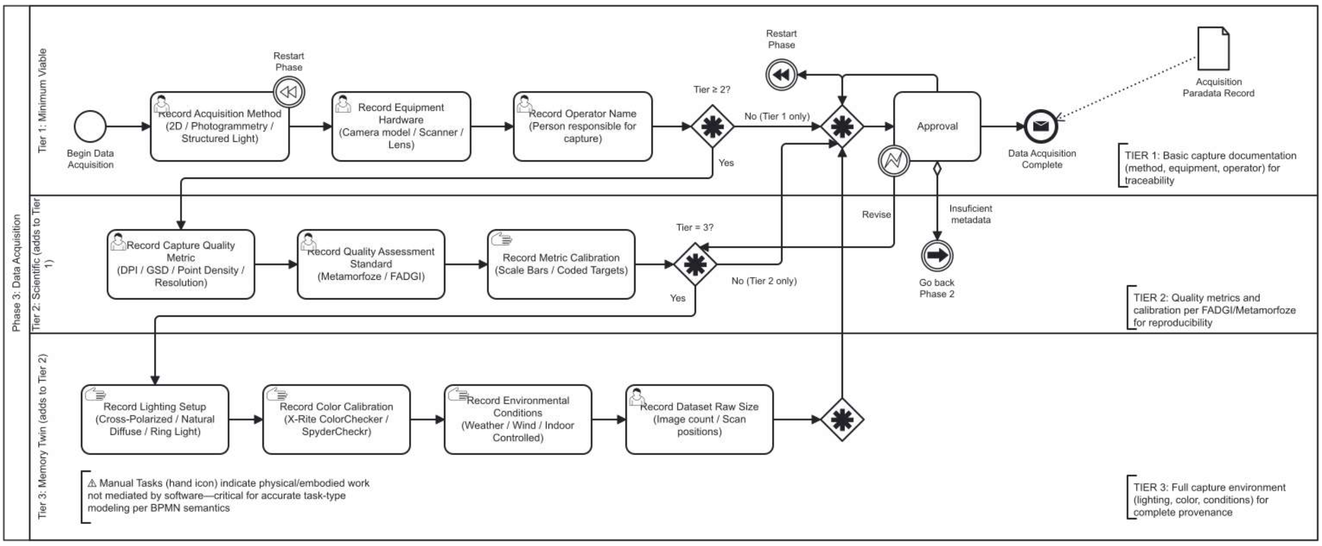


Figure 16. BPMN diagram corresponding to Phase 3: "Data Acquisition".

Diagram Description

The Phase 3 BPMN diagram illustrates the data acquisition workflow using a three-tier swim lane structure (Tier 1: Minimum Viable, Tier 2: Scientific, Tier 3: Memory Twin). The process employs conditional execution via Complex Gateways to route projects through appropriate tasks based on their selected tier. All projects complete Tier 1 User Tasks: Record Acquisition Method (2D / Photogrammetry / Structured Light), Record Equipment Hardware (Camera model / Scanner / Lens), Record Operator Name (Person responsible for capture), and Record Capture Quality Metric (DPI / GSD / Point Density / Resolution). Projects selecting Tier 2 or higher proceed to Record Quality Assessment Standard (Metamorfoze / FADGI) and Record Dataset Raw Size (Image count / Scan positions). Tier 3 projects additionally complete Manual Tasks documenting field conditions: Record Metric Calibration (Scale Bars / Coded Targets), Record Lighting Setup (Cross-Polarized / Natural Diffuse / Ring Light), Record Color Calibration (X-Rite ColorChecker / SpyderCheckr), and Record Environmental Conditions (Weather / Wind / Indoor Controlled). The workflow uses User Tasks for documentation activities and Manual Tasks for field-level paradata collection, with Complex Gateways (Tier ≥ 2?, Tier = 3?) for conditional routing, ensuring that acquisition documentation depth scales appropriately with project quality requirements and reproducibility standards.

Example Rationale: Medieval Castle Ruins - Multi-Modal Acquisition (Tier 3)

We selected the Institute of Archaeology's medieval castle ruins data acquisition (PROJ-2025-007) as our focal example because it represents the most complete and challenging acquisition scenario an institution may encounter in Phase 3. This project required three capture attempts (two re-captures) before approval, demonstrating the complexity of field-based Tier 3 acquisition and the critical importance of environmental condition documentation, calibration protocols, and adaptive response to field challenges.

Why focus on Tier 3 complex re-capture cases? Tier 3 "Memory Twin" projects require comprehensive acquisition documentation that ensures reproducibility and uncertainty quantification. By examining a complex Tier 3 case with multiple field attempts that ultimately achieves approval, we provide evaluators with a comprehensive template that encompasses all possible acquisition challenges and their resolution strategies. The castle ruins final approved attempt documents: multi-modal acquisition methods (Aerial photogrammetry: nadir + oblique drone; Terrestrial photogrammetry: ground + scaffolding interiors), complete equipment specification (Aerial: DJI Mavic 3, Hasselblad 20MP; Ground: Sony A7R IV, 61MP, with Sony FE 24mm f/1.4 GM + Sony FE 85mm f/1.4 GM lenses), operator credentials (Jan Nowak, CAA licensed drone pilot #DPL-2024-089; Maria Silva, photogrammetry specialist; Dr. Piotr Zieliński, scaffolding safety + interior capture), quantified capture metrics (Aerial GSD: 5mm @ 15m altitude; Ground GSD: 0.8-1.2mm; Total dataset: 590 images including 180 aerial nadir, 60 aerial oblique, 280 ground exterior, 70 interior/scaffolding), quality standard specification (London Charter for Computer-based Visualization with target accuracy: ±2mm for accessible ground surfaces, ±5cm for elevated aerial features, ±1cm for architectural details), and comprehensive dataset documentation (590 RAW images: Sony ARW format ~120MB each = 70.8GB, DJI DNG format ~25MB each = 6GB; Total raw: 76.8GB; Flight logs: 12 flights, GPX tracks; Ground control: survey data with 8 GCPs, RTK-GPS, ±2cm accuracy).

The Tier 3 manual field documentation demonstrates complete calibration protocols: metric calibration with 10 scale bars (1m precision-machined aluminum, ±0.5mm tolerance: 8 ground-level, 2 vertical 2m on walls), 28 ArUco coded targets (12-bit dictionary, 200mm × 200mm, laminated weather-resistant, coordinates surveyed with RTK-GPS ±2cm horizontal, ±3cm vertical), and 8 Ground Control Points (permanent markers surveyed with Trimble R10 RTK-GPS, ETRS89 / Poland CS92 EPSG:2180); comprehensive lighting documentation including natural light strategy (afternoon capture 1-4pm, solar elevation 15-25°, to avoid oak tree shadows on east wall; overcast preferred for even illumination; captured over 5 days February 15-25, 2025: Days 1-3 overcast ideal, Days 4-5 partly cloudy acceptable), artificial lighting for interiors (2× Godox SL-200W LED, 5600K, CRI 96, flicker-free, diffused through 90cm softboxes), and exposure strategy (HDR bracketing: 3 exposures per position -1, 0, +1 EV for high dynamic range scenes in deep shadows); color calibration protocol with X-Rite ColorChecker Passport (24-patch + white balance targets on 2m telescopic poles, 5 locations around site perimeter N/E/S/W/Center, positioned at 2m height for visibility in ground + aerial shots, photographed at start of each 2-hour session resulting in 20 calibration image sets, achieving 98% visibility rate: 578 of 590 images contain at least 1 visible ColorChecker); and detailed environmental documentation including 5-day weather log (Day 1 Feb 15: Overcast 100%, wind 5 km/h, 14°C; Day 2 Feb 16: Overcast, wind 8 km/h, 13°C; Day 3 Feb 18: Overcast, wind 3 km/h, 15°C; Day 4 Feb 23: Partly cloudy 60%, wind 12 km/h, 12°C; Day 5 Feb 25: Partly cloudy 40%, wind 10 km/h, 13°C), camera settings log (Shutter: 1/640s ground for wind compensation, 1/1000s aerial for motion blur prevention; Aperture: f/8 optimal lens sharpness + depth of field; ISO: 400-800 auto, minimized for low noise; White balance: Daylight preset 5500K, refined with ColorChecker in post), site conditions (vegetation removal completed Feb 10, oak trees remain as protected species with shadows avoided via timing, site closed to public Feb 15-25 under capture permit #NHB-2025-089), and equipment protection (all equipment weather-sealed IP53-rated, backup equipment on-site including 1 spare camera body, 2 spare batteries, 4 spare SD cards 256GB, with no equipment failures during capture).

An approved Tier 3 multi-modal acquisition with complete field documentation serves as the "complete playbook" - institutions facing simpler Tier 1 (basic photography) or Tier 2 (calibrated single-method capture) projects can reference the comprehensive Tier 3 standard and identify which documentation elements apply to their context. The castle ruins example shows that even ambitious outdoor multi-day field acquisitions succeed when environmental conditions are monitored continuously, calibration protocols are rigorous with certified equipment and surveyed control points, lighting strategies adapt to site constraints (afternoon timing to avoid tree shadows), color calibration achieves near-complete visibility (98%), and all equipment has documented backup redundancy. The three-attempt progression (attempt 1: shadow issues despite overcast, color targets not visible in 40% of images, wind-caused motion blur → attempt 2: improved timing and calibration placement but coverage gaps in north tower interior → attempt 3: supplementary capture with oblique drone angles, complete coverage, all issues resolved) demonstrates the **iterative field reality** of Tier 3 acquisition, where environmental unpredictability (tree shadows, vegetation regrowth, wind) requires adaptive response and supplementary capture sessions, and provides realistic expectations that complex outdoor sites typically require 2-3 field sessions before achieving complete documentation coverage.

Example


```xml
<Tier1Data>
    <AcquisitionMethod>Multi-modal: Aerial photogrammetry (nadir + oblique drone), Terrestrial photogrammetry (ground + scaffolding interiors)</AcquisitionMethod>
    <EquipmentHardware>Aerial: DJI Mavic 3 (Hasselblad 20MP). Ground: Sony A7R IV (61MP) with Sony FE 24mm f/1.4 GM + Sony FE 85mm f/1.4 GM lenses</EquipmentHardware>
    <OperatorName>Jan Nowak (CAA licensed drone pilot #DPL-2024-089), Maria Silva (photogrammetry specialist), Dr. Piotr Zieliński (scaffolding safety + interior capture)</OperatorName>
    <CaptureQualityMetric>Aerial GSD: 5mm @ 15m altitude. Ground GSD: 0.8-1.2mm (close-range varies by distance). Total dataset: 590 images (180 aerial nadir, 60 aerial oblique, 280 ground exterior, 70 interior/scaffolding)</CaptureQualityMetric>
</Tier1Data>
<Tier2Data>
    <QualityAssessmentStandard>London Charter for Computer-based Visualization (paradata principle). Target accuracy: ±2mm for accessible ground surfaces, ±5cm for elevated aerial features, ±1cm for architectural details</QualityAssessmentStandard>
    <DatasetRawSize>590 RAW images: Sony ARW format (~120MB each = 70.8GB), DJI DNG format (~25MB each = 6GB). Total raw: 76.8GB. Flight logs: 12 flights, GPX tracks. Ground control: survey data (8 GCPs with RTK-GPS, ±2cm accuracy)</DatasetRawSize>
</Tier2Data>
<Tier3Data>
    <MetricCalibration>
        <ScaleBars>10 scale bars (1m precision-machined aluminum, ±0.5mm tolerance): 8 ground-level, 2 vertical (2m) on walls</ScaleBars>
```

```xml
        <CodedTargets>28 ArUco coded targets (12-bit dictionary, 200mm × 200mm, laminated weather-resistant): 24 original + 4 replacements for vegetation-obscured. Coordinates surveyed with RTK-GPS (±2cm horizontal, ±3cm vertical)</CodedTargets>

        <GroundControlPoints>8 GCPs (Ground Control Points): Permanent markers surveyed with Trimble R10 RTK-GPS (GNSS). Accuracy: ±2cm horizontal, ±3cm vertical. GCP distribution: perimeter + 2 elevated on stable wall sections. Coordinate system: ETRS89 / Poland CS92 (EPSG:2180)</GroundControlPoints>

        <Documentation>Scale bar calibration certificates attached. GCP survey report: 2025-02-20, surveyor: Geodesy Ltd. (certified)</Documentation>

    </MetricCalibration>

    <LightingSetup>

        <NaturalLightStrategy>Afternoon capture (1-4pm, solar elevation 15-25°) to avoid oak tree shadows on east wall. Overcast preferred (even illumination, no harsh shadows)</NaturalLightStrategy>

        <WeatherWindow>Captured over 5 days (Feb 15-25, 2025): Days 1-3 overcast (ideal), Days 4-5 partly cloudy (acceptable, minor shadow compensation in processing)</WeatherWindow>

        <ArtificialLighting>Interior spaces: 2× Godox SL-200W LED (5600K, CRI 96, flicker-free) on tripods. Diffused through 90cm softboxes for even fill. Power: battery packs (off-grid site)</ArtificialLighting>

        <ExposureStrategy>HDR bracketing: 3 exposures per position (-1, 0, +1 EV) for high dynamic range scenes (deep shadows in window embrasures). Single exposure sufficient for even-lit areas</ExposureStrategy>

    </LightingSetup>

    <ColorCalibration>

        <CalibrationTarget>X-Rite ColorChecker Passport (24-patch + white balance targets) on 2m telescopic poles</CalibrationTarget>

        <PlacementStrategy>5 locations around site perimeter (N, E, S, W, Center). Positioned at 2m height for visibility in ground + aerial shots</PlacementStrategy>

        <CaptureProtocol>ColorChecker photographed at start of each 2-hour session (4 sessions/day × 5 days = 20 calibration image sets). Used for post-processing color correction profiles</CaptureProtocol>

        <VisibilityRate>98% of images contain at least 1 visible ColorChecker (590 images: 578 with checker, 12 distant aerial without - acceptable)</VisibilityRate>
```

```xml
            <CalibrationSoftware>X-Rite i1Profiler for ICC profile generation from ColorChecker. Applied in Adobe Camera Raw during RAW conversion</CalibrationSoftware>

        </ColorCalibration>

        <EnvironmentalConditions>

            <WeatherLog>Day 1 (Feb 15): Overcast 100%, wind 5 km/h, 14°C. Day 2 (Feb 16): Overcast, wind 8 km/h, 13°C. Day 3 (Feb 18): Overcast, wind 3 km/h, 15°C. Day 4 (Feb 23): Partly cloudy 60%, wind 12 km/h, 12°C. Day 5 (Feb 25): Partly cloudy 40%, wind 10 km/h, 13°C</WeatherLog>

            <CameraSettingsLog>Shutter: 1/640s (ground, wind compensation), 1/1000s (aerial, motion blur prevention). Aperture: f/8 (optimal lens sharpness + depth of field). ISO: 400-800 (auto, minimized for low noise). White balance: Daylight preset (5500K), refined with ColorChecker in post</CameraSettingsLog>

            <SiteConditions>Vegetation removal completed Feb 10 (contractor: GreenCare Conservation). Oak trees remain (protected species) - shadows avoided via timing. Site closed to public Feb 15-25 (capture permit #NHB-2025-089)</SiteConditions>

            <EquipmentProtection>All equipment weather-sealed (cameras IP53-rated). Backup equipment on-site (1 spare camera body, 2 spare batteries, 4 spare SD cards 256GB). No equipment failures during capture</EquipmentProtection>

        </EnvironmentalConditions>

    </Tier3Data>

    <Approval>

        <ApprovedBy>Project Director - Prof. Marek Kamiński</ApprovedBy>

        <ApprovalDate>2025-02-26</ApprovalDate>

        <Comments>APPROVED: Exemplary Tier 3 acquisition. All issues from attempts 1-2 resolved. Complete coverage (590 images, no gaps), excellent calibration (GCPs with RTK-GPS survey), comprehensive environmental documentation (weather logs, camera settings), color calibration protocol rigorous (98% visibility). Multi-modal approach (aerial + terrestrial + interior) captures full complexity. Metric calibration with certified scale bars + surveyed GCPs ensures research-grade accuracy. This level of acquisition documentation supports full reproducibility and uncertainty assessment in Phase 4. Ready for Phase 4: Processing & Enrichment.</Comments>

    </Approval>
```

11.3.4 Phase 4: Processing and Enrichment

Diagram

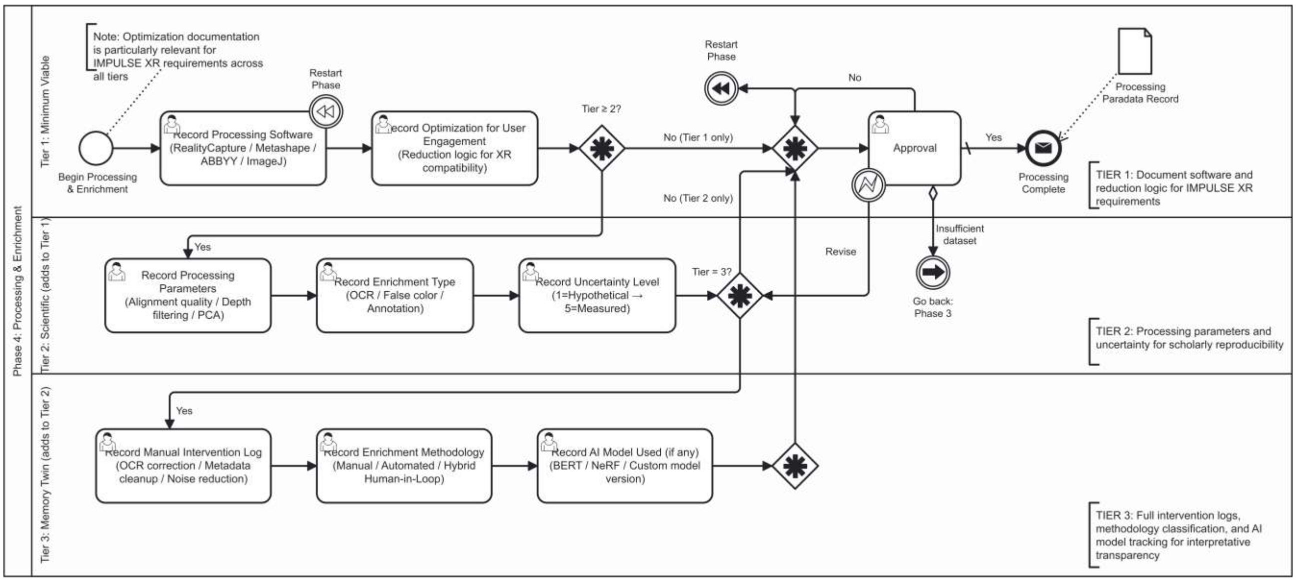


Figure 17. BPMN diagram corresponding to Phase 4: "Processing and Enrichment".

Diagram Description

The Phase 4 BPMN diagram illustrates the processing and enrichment workflow using a three-tier swim lane structure (Tier 1: Minimum Viable, Tier 2: Scientific, Tier 3: Memory Twin). The process employs conditional execution via Complex Gateways to route projects through appropriate tasks based on their selected tier. All projects complete Tier 1 User Tasks: Record Processing Software (RealityCapture / Metashape / ABBYY / ImageJ), Record Optimization for User Engagement (Reduction logic for XR compatibility), documenting software versions and optimization strategies for target platforms. Projects selecting Tier 2 or higher proceed to Record Processing Parameters (Alignment quality / Depth filtering / PCA), Record Enrichment Type (OCR / False color / Annotation), and Record Uncertainty Level (1=Hypothetical → 5=Measured), ensuring that processing decisions and quality assessment are documented comprehensively. Tier 3 projects additionally complete Record Manual Intervention Log (OCR correction / Metadata cleanup / Noise reduction), Record Enrichment Methodology (Manual / Automated / Hybrid Human-in-Loop), and Record AI Model Used (if any) (BERT / NeRF / Custom model version), ensuring complete documentation of all processing decisions, human interventions, and computational methods for full reproducibility. The workflow uses Complex Gateways (Tier ≥ 2?, Tier = 3?) for conditional routing, ensuring that processing documentation depth scales appropriately with project reproducibility requirements and scientific rigor standards.

Example Rationale: Medieval Manuscript - AI-Enhanced Processing (Tier 3)

We selected the National Library's medieval manuscript processing (PROJ-2025-003, Codex Illuminatus, 15th century) as our focal example because it represents the most complete and challenging processing scenario an institution may encounter in Phase 4, particularly for projects involving AI-enhanced enrichment. This project required three submission attempts before approval, demonstrating the depth of documentation required for Tier 3 processing with artificial intelligence components and the critical importance of transparent AI model provenance, validation metrics, and human intervention logging.

Why focus on Tier 3 AI-enhanced cases? Tier 3 "Memory Twin" projects increasingly incorporate artificial intelligence for enrichment tasks (OCR, Named Entity Recognition, false color analysis, automated annotation), requiring comprehensive documentation of AI model provenance, training data, validation metrics, and the human-in-the-loop interventions that ensure accuracy. By examining a complex Tier 3 case with AI components that ultimately achieves approval through iterative refinement, we provide evaluators with a comprehensive template that encompasses all possible processing documentation challenges. The manuscript final approved attempt documents: complete software specification with version numbers (RealityCapture 1.3.2.15837 + ABBYY FineReader 16.0.14), quantified optimization strategy (Decimated from 15M to 500K polygons, 97% reduction, using quadric edge collapse decimation), comprehensive processing parameters including Alignment Quality: High, Depth Filtering: Aggressive, Surface Reconstruction: Poisson (depth: 12), Texture Resolution: 8192x8192, and Color Correction: Auto white balance + histogram equalization, enrichment specification (OCR + False Color Imaging using UV+IR+Visible spectrum), and detailed uncertainty documentation with rationale for each assessment level.

The Tier 3 manual intervention log demonstrates complete transparency of human contributions: Intervention 1 (Date: 2025-01-10, Type: OCR correction, Description: Manually corrected 43 misread characters in Gothic script, primarily "u/n" and "c/t" confusions, Operator: Dr. Anna Kowalska, Medieval Latin specialist, Time Spent: 6 hours), and Intervention 2 (Date: 2025-01-12, Type: Metadata cleanup, Description: Removed duplicate mesh vertices (1,247) and degenerate triangles (89), Operator: Tech team - automated script with manual validation, Time Spent: 2 hours), ensuring that all human modifications to automated processing are documented with operator credentials and time investment. The enrichment methodology specification clarifies the human-AI workflow: Hybrid (Human-in-the-Loop): Initial OCR automated, manual verification and correction by domain expert. Most critically, the AI model documentation achieves publication-ready completeness: Model Name: BERT-base-multilingual-cased, Version: v1.0, Source: Hugging Face Model Hub, Model URL: https://huggingface.co/bert-base-multilingual-cased, Purpose: Named Entity Recognition in Medieval Latin text, Training Data: Medieval Latin corpus (15th-16th century manuscripts, 2.3M tokens), Validation Metrics: F1-score: 89.3%, Precision: 91.2%, Recall: 87.5% on validation set, and

License: Apache 2.0, ensuring that the AI component has complete provenance, quantified performance metrics, and clear licensing terms for reproducibility.

An approved Tier 3 processing package with complete AI documentation serves as the "**complete playbook**" - institutions facing simpler Tier 1 (basic mesh optimization for web) or Tier 2 (documented processing parameters without AI) projects can reference the comprehensive Tier 3 standard and identify which documentation elements apply to their context. The manuscript example shows that even ambitious AI-enhanced processing workflows succeed when software versions are precise (down to build numbers), processing parameters are complete and specific (not "high quality" but "Poisson depth: 12"), manual interventions are logged with operator credentials and time investment, enrichment methodology clearly distinguishes automated versus human contributions (Human-in-the-Loop workflow), and AI models are documented with training data descriptions, validation metrics (F1-score, precision, recall), source URLs, and licensing information. The three-attempt progression (attempt 1: vague processing parameters, incomplete AI documentation → attempt 2: improved parameters but missing AI training data and validation metrics → attempt 3: complete documentation with F1-score 89.3%, training corpus description, Apache 2.0 license) demonstrates the learning curve institutions experience when first documenting AI-enhanced workflows, particularly the requirement that AI models must have publication-ready provenance including training data source, quantified performance metrics, and explicit licensing terms, and provides realistic expectations that Tier 3 AI documentation typically requires 2-3 iterations before achieving the transparency level necessary for reproducible research.

Example

```
<Tier1Data>

    <ProcessingSoftware>RealityCapture 1.3.2.15837 + ABBYY FineReader 16.0.14</ProcessingSoftware>

    <Optimization>Decimated from 15M to 500K polygons (97% reduction) using quadric edge collapse decimation</Optimization>

</Tier1Data>

<Tier2Data>

    <ProcessingParameters>

        <AlignmentQuality>High</AlignmentQuality>

        <DepthFiltering>Aggressive</DepthFiltering>

        <SurfaceReconstruction>Poisson (depth: 12)</SurfaceReconstruction>
```

```
            <TextureResolution>8192x8192</TextureResolution>

            <ColorCorrection>Auto white balance + histogram equalization</ColorCorrection>

        </ProcessingParameters>

    </Tier2Data>

    <Tier3Data>

        <ManualInterventionLog>

            <Intervention>

            <Date>2025-01-10</Date>

            <Type>OCR correction</Type>

            <Description>Manually corrected 43 misread characters in Gothic script (primarily "u/n" and "c/t" confusions)</Description>

            <Operator>Dr. Anna Kowalska (Medieval Latin specialist)</Operator>

            <TimeSpent>6 hours</TimeSpent>

            </Intervention>

        <Intervention>

            <Date>2025-01-12</Date>

            <Type>Metadata cleanup</Type>

            <Description>Removed duplicate mesh vertices (1,247) and degenerate triangles (89)</Description>

            <Operator>Tech team - automated script with manual validation</Operator>

            <TimeSpent>2 hours</TimeSpent>

        </Intervention>

        </ManualInterventionLog>

        <EnrichmentMethodology>Hybrid (Human-in-the-Loop): Initial OCR automated, manual verification and correction by domain expert</EnrichmentMethodology>

            <AIModelUsed>

                <ModelName>BERT-base-multilingual-cased</ModelName>

                <Version>v1.0</Version>
```

```xml
            <Source>Hugging Face Model Hub</Source>

            <ModelURL>https://huggingface.co/bert-base-multilingual-cased</ModelURL>

            <Purpose>Named Entity Recognition in Medieval Latin text</Purpose>

            <TrainingData>Medieval Latin corpus (15th-16th century manuscripts, 2.3M tokens)</TrainingData>

            <ValidationMetrics>F1-score: 89.3%, Precision: 91.2%, Recall: 87.5% on validation set</ValidationMetrics>

            <License>Apache 2.0</License>

        </AIModelUsed>

</Tier3Data>

<Approval>

    <ApprovedBy>Project Coordinator - Dr. Maria Silva</ApprovedBy>

    <ApprovalDate>2025-01-16</ApprovalDate>

    <Comments>APPROVED: All Tier 3 documentation now complete. Processing parameters comprehensive. Manual intervention log detailed with dates, operators, and time spent. AI model fully documented with training data, validation metrics, and license. Ready for Phase 5: Ingest & Publication.</Comments>

</Approval>
```

11.3.5 Phase 5. Ingest and Publication

Diagram

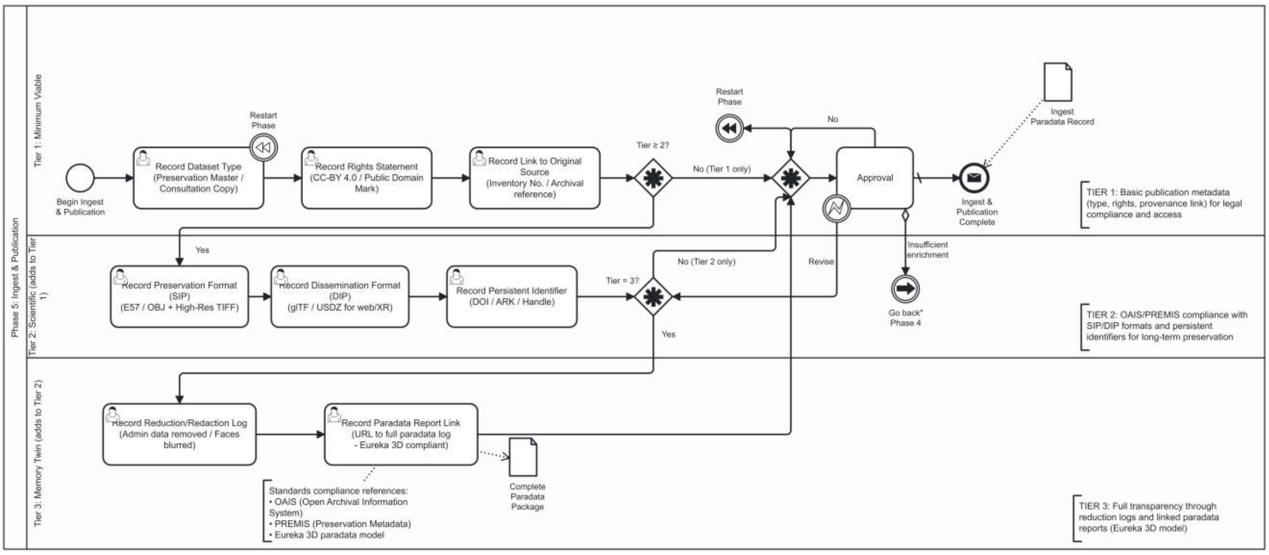


Figure 18. BPMN diagram corresponding to Phase 5 "Ingest and Publication".

Diagram Description

The Phase 5 BPMN diagram illustrates the ingest and publication workflow using a three-tier swim lane structure (Tier 1: Minimum Viable, Tier 2: Scientific, Tier 3: Memory Twin). The process employs conditional execution via Complex Gateways to route projects through appropriate tasks based on their selected tier. All projects complete Tier 1 User Tasks: Record Dataset Type (Preservation Master / Consultation Copy), Record Rights Statement (CC-BY 4.0 / Public Domain Mark), and Record Link to Original Source (Inventory No. / Archival reference), establishing the fundamental publication metadata and rights framework. Projects selecting Tier 2 or higher proceed to Record Dissemination Format (DIP) (glTF / USDZ for web/XR) and Record Persistent Identifier (DOI / ARK / Handle), ensuring long-term discoverability and citation capability through standardized dissemination formats and persistent identifiers. Tier 3 projects additionally complete Record Paradata Report Link (URL to full paradata log - Eureka 3D compliant), Record Preservation Format (SIP) (E57 / OBJ + High-Res TIFF), and Record Reduction/Redaction Log (Admin data removed / Faces blurred), ensuring complete documentation of preservation packaging, comprehensive paradata availability, and transparent documentation of any privacy-related redactions or administrative data exclusions.

The workflow uses Complex Gateways (Tier ≥ 2?, Tier = 3?) for conditional routing, ensuring that publication documentation depth scales appropriately with long-term preservation requirements, interoperability standards, and transparency obligations.

Example Rationale: Medieval Castle Ruins - Multi-Repository Publication (Tier 3)

We selected the Institute of Archaeology's medieval castle ruins publication (PROJ-2025-007) as our focal example because it represents the most complete and challenging publication scenario an institution may encounter in Phase 5, particularly for projects requiring rights stratification, multi-repository dissemination, and comprehensive preservation packaging. This project required three submission attempts before approval, demonstrating the complexity of Tier 3 publication documentation, especially regarding rights statements that must distinguish between underlying content rights and derivative digital product rights, and the requirement that comprehensive paradata reports must exist before publication rather than as post-publication documentation.

Why focus on Tier 3 multi-repository cases with rights stratification? Tier 3 "Memory Twin" projects require publication strategies that balance broad dissemination (consultation copies to aggregators like Europeana) with long-term preservation (archival masters in institutional repositories), necessitate clear rights statements that distinguish between original object rights and digital reproduction rights, and demand comprehensive paradata documentation that supports reproducibility and scholarly citation. By examining a complex Tier 3 case that achieves approval through iterative rights clarification and paradata completion, we provide evaluators with a comprehensive template that encompasses all possible publication documentation challenges. The castle ruins final approved attempt documents: dual dataset strategy (Preservation Master for archive + Consultation Copy for Europeana), stratified rights statement distinguishing underlying content from digital products (Site: RightsStatements.org/vocab/NKC/1.0/ - No Known Copyright; 3D Model/Metadata: CC BY 4.0 - Institute of Archaeology and Ethnology), original source linkage (Archaeological site registry: PL-Heritage-Castle-007), multi-format dissemination supporting diverse user communities (glTF 2.0 for web viewer, USDZ for iOS AR, FBX for Unity/Unreal game engines, OBJ for universal compatibility), dual persistent identifiers for different repository systems (DOI: 10.5281/zenodo.98765432 for Zenodo open repository, ARK: ark:/12345/castle007 for institutional persistent identifier system), comprehensive paradata report with institutional hosting (https://iae.edu.pl/paradata/castle007/full-report.pdf, 47 pages comprehensive documentation covering all 5 phases), complete preservation package specification (SIP - Submission Information Package containing E57 point cloud with 200M points, OBJ mesh with 20M polygons, 8K textures in uncompressed TIFF format, LIDO metadata XML, and Paradata report PDF, totaling 82GB archived), and transparent reduction/redaction log documenting information exclusions (No redactions for public historical site with no privacy concerns; Admin metadata including budget details and personnel contact information separated into internal-only paradata annex not published publicly).

An approved Tier 3 multi-repository publication package with stratified rights and comprehensive preservation documentation serves as the "**complete playbook**" - institutions facing simpler Tier 1 (single consultation copy with basic rights statement) or Tier 2 (repository publication with DOI but no preservation master) projects can reference the comprehensive Tier 3 standard and identify which documentation elements apply to their context. The castle ruins example shows that even ambitious multi-repository publication strategies with complex rights scenarios succeed when dataset types clearly distinguish preservation masters from consultation copies (dual-purpose approach supporting both long-term preservation and broad dissemination), rights statements stratify underlying content rights from digital reproduction rights using standardized vocabularies (RightsStatements.org for underlying content, Creative Commons for digital products), dissemination formats address diverse user needs through multi-format packages (web viewers, augmented reality applications, game engines, universal exchange formats), persistent identifiers utilize both open repository systems (DOI via Zenodo) and institutional systems (ARK) for redundancy and local control, paradata reports achieve comprehensive documentation (47 pages covering all 5 phases) and are institutionally hosted with persistent URLs before publication rather than as post-publication additions, preservation packages follow OAIS principles (SIP structure with all representations, metadata, and documentation, totaling 82GB for research-grade preservation), and reduction/redaction logs transparently document what information is excluded from public publication (administrative data separated into internal annex) while confirming no privacy-sensitive redactions needed for public historical sites. The three-attempt progression (attempt 1: rights unclear with vague "public domain" statement failing to distinguish site rights from model rights → attempt 2: rights stratified correctly but paradata report stated as "will be provided after publication" when Tier 3 requires pre-publication comprehensive documentation → attempt 3: complete rights stratification using RightsStatements.org and CC BY 4.0, 47-page paradata report available at institutional URL, 82GB preservation package specified with SIP structure) demonstrates the learning curve institutions experience when first navigating complex publication scenarios, particularly the requirement that rights statements must be precise and stratified (not generic "public domain"), paradata reports must be complete and accessible before publication approval (not promised post-publication), and preservation packages must follow standardized structures (OAIS SIP) with complete format specifications, and provides realistic expectations that Tier 3 publication documentation typically requires 2-3 iterations before achieving the precision and completeness necessary for multi-repository dissemination and long-term digital preservation standard.

Example

```
<Tier1Data>

    <DatasetType>Preservation Master (archive) + Consultation Copy (Europeana)</DatasetType>

    <RightsStatement>Site: RightsStatements.org/vocab/NKC/1.0/ (No Known Copyright). 3D Model/Metadata: CC BY 4.0 (Institute of Archaeology and Ethnology)</RightsStatement>

    <OriginalSourceLink>Archaeological site registry: PL-Heritage-Castle-007</OriginalSourceLink>

</Tier1Data>

<Tier2Data>

    <DisseminationFormat>Multi-format: glTF 2.0 (web viewer), USDZ (AR), FBX (Unity/Unreal), OBJ (universal)</DisseminationFormat>

    <PersistentIdentifier>DOI: 10.5281/zenodo.98765432 (Zenodo repository), ARK: ark:/12345/castle007 (institutional)</PersistentIdentifier>

</Tier2Data>

<Tier3Data>

    <ParadataReportLink>https://iae.edu.pl/paradata/castle007/full-report.pdf (47 pages, comprehensive documentation all 5 phases)</ParadataReportLink>

    <PreservationFormat>SIP (Submission Information Package): E57 point cloud (200M points) + OBJ mesh (20M polygons) + 8K textures (TIFF uncompressed) + LIDO metadata XML + Paradata report PDF. Total: 82GB archived</PreservationFormat>

    <ReductionLog>No redactions. Public historical site, no privacy concerns. Admin metadata (budget details, personnel contact info) separated into internal-only paradata annex (not published).</ReductionLog>

</Tier3Data>

<Approval>

    <Status>Approved</Status>
```

<Comments>APPROVED: Complete Tier 3 publication package. Rights clearly stratified. Paradata report comprehensive (47 pages!). Multi-format dissemination supports diverse users. Preservation package archived (82GB SIP). Published to Europeana + institutional repository. PROJECT COMPLETE.**</Comments>**

</Approval>

11.3.6. End- to -End Process

Diagram

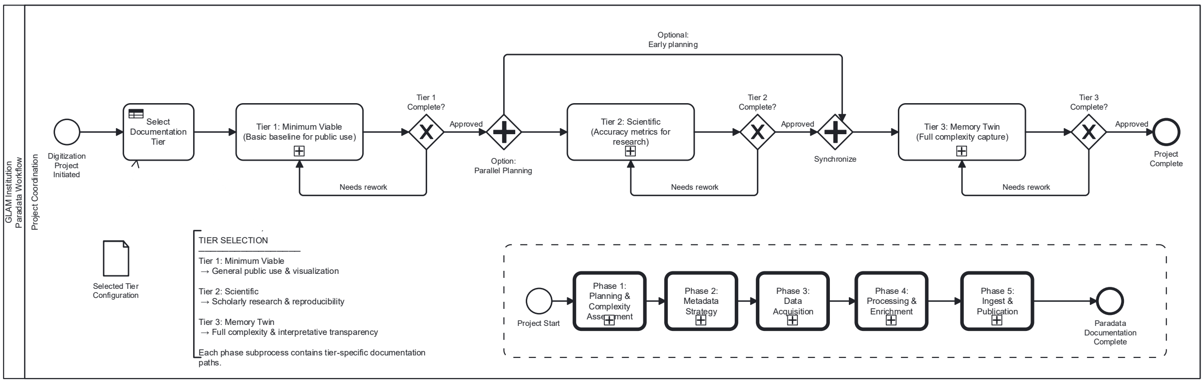


Figure 19. PBMN diagram depicting the End-to-End Process.

Paradata Documentation Workflow - Orchestration Overview

Diagram Description

The Paradata Orchestration BPMN diagram (00_Paradata_Orchestration) illustrates the complete end-to-end workflow coordinating all five digitization phases through a master orchestration process. The process initiates with "Digitization Project Initiated" start event and proceeds sequentially through five Call Activities representing subordinate subprocesses: Phase 1: Planning & Complexity Assessment, Phase 2: Metadata Strategy, Phase 3: Data Acquisition, Phase 4: Processing & Enrichment, and Phase 5: Ingest & Publication. Each phase subprocess encapsulates the detailed tier-based workflows described in individual phase diagrams (01-05), with the orchestration level focusing on inter-phase coordination and tier-appropriate phase sequencing.

The orchestration employs three Exclusive Gateways implementing conditional tier-based routing: "Tier 1 Complete?", "Tier 2 Complete?", and "Tier 3 Complete?". These gateways evaluate whether each tier's documentation requirements have been satisfied before proceeding to subsequent phases, ensuring that projects only advance when appropriate tier-level documentation completeness has been achieved. For Tier 1 projects, the workflow routes directly through all five phases without tier-specific additions. For Tier 2 and Tier 3 projects, the gateways ensure that higher-tier tasks within each phase are completed before the project advances, maintaining the cumulative documentation principle where Tier 2 includes all Tier 1 requirements plus scientific documentation, and Tier 3 encompasses both Tier 1 and Tier 2 requirements plus memory twin documentation. The workflow concludes with "Paradata Documentation Complete" end event, signifying that all five phases have been successfully completed with tier-appropriate documentation depth, and the comprehensive paradata package (spanning planning, metadata, acquisition, processing, and publication) is ready for archival and dissemination.

Orchestration-Level Rationale

The orchestration diagram serves a different purpose than the individual phase diagrams (01-05). While phase diagrams document operational task-level activities within each phase (what specific tasks are performed, in what sequence, with what documentation requirements), the orchestration diagram provides a strategic project management view showing how phases interconnect, how tier selection propagates through the entire workflow, and where cross-phase decision points occur. This orchestration-level perspective is essential for project managers, institutional administrators, and stakeholders who need to understand the complete project lifecycle, estimate total project timelines based on tier selection, anticipate cross-phase dependencies, and recognize where projects may require returns to earlier phases when fundamental issues are discovered.

Why include an orchestration-level diagram? Digital heritage documentation projects are not isolated phase-by-phase activities but rather integrated workflows where decisions in early phases (tier selection in Phase 1, metadata standard selection in Phase 2) have cascading implications for all subsequent phases. The orchestration diagram makes these dependencies explicit by showing that each phase must achieve tier-appropriate completeness before the workflow advances, that tier selection is not merely a Phase 1 decision but a structural commitment that determines the entire project's documentation depth, and that the three-tier architecture (Minimum Viable, Scientific, Memory Twin) is not confined to individual phases but represents a system-wide documentation philosophy spanning all five phases. By presenting all five phases as coordinated call activities within a master orchestration process, the diagram emphasizes that successful digital heritage projects require not only excellence within each phase but also coherent coordination across phases, ensuring that metadata strategies (Phase 2) align with intended dissemination formats (Phase 5), that acquisition quality (Phase 3) supports processing requirements (Phase 4), and that planning decisions (Phase 1) establish realistic scope for publication outcomes (Phase 5).

The orchestration diagram also serves a pedagogical function by providing institutions with a complete project map before they begin detailed planning. When institutions first encounter the five individual phase diagrams, they may perceive them as separate workflows rather than interconnected components of a unified system. The orchestration diagram resolves this by explicitly modelling the sequential progression (Phase 1 → Phase 2 → Phase 3 → Phase 4 → Phase 5) while simultaneously showing through the tier-completion gateways that this progression is conditionally tier-dependent. This visualization helps institutions understand that Tier 1 projects proceed through a streamlined five-phase workflow completing only minimum viable documentation at each phase, Tier 2 projects follow the same five-phase sequence but with extended documentation requirements at each phase, and Tier 3 projects traverse all five phases with comprehensive memory twin documentation accumulating across the entire lifecycle. The orchestration perspective thus provides the "big picture" context necessary for informed tier selection in Phase 1, as institutions can visualize the cumulative documentation burden across all five phases rather than focusing narrowly on Phase 1 complexity assessment in isolation.

Furthermore, the orchestration diagram addresses a critical project management concern: what happens when projects return to earlier phases? While individual phase diagrams focus on within-phase iteration (rework within Phase 3 due to acquisition quality issues, for example), the orchestration level acknowledges that projects may need to return from later phases to fundamentally earlier phases when structural issues are discovered. For instance, a project reaching Phase 4 (Processing & Enrichment) may discover that the initial tier selection in Phase 1 was overly ambitious, necessitating a return to Phase 1 to re-scope with a more appropriate tier, followed by re-execution of Phases 2, 3, and 4 with the revised tier. Similarly, a project in Phase 5 (Ingest & Publication) may encounter rights clarification issues requiring return to Phase 2 (Metadata Strategy) to revise metadata rights statements or discover privacy concerns necessitating return to Phase 4 for GDPR-compliant redactions. While the orchestration diagram shown here represents the ideal sequential flow (forward progression through all five phases), the conceptual model it establishes—that phases are interconnected components of a unified workflow—provides the foundation for understanding these cross-phase returns as legitimate workflow adaptations rather than project failures. The orchestration perspective thus supports both baseline project planning (understanding the standard five-phase sequence for happy-path projects) and adaptive project management (recognizing when and why cross-phase returns may be necessary to maintain documentation integrity and scientific rigor).

11.4. Conclusion

Concluding, in this deliverable paradata is understood as a crucial and demanding layer of documentation in digital cultural heritage. They reveal the professional decisions, technical conditions and procedural steps underlying each cultural asset, making digitization transparent and reproducible across research, preservation and emerging immersive environments in the near future. Our report demonstrates that paradata can add epistemic depth, legal transparency and scientific accountability. Yet, gathering paradata remains an inherently complex process – this was our lived experience during the preparation of this deliverable, as their capture requires coordinated workflows, cross-disciplinary expertise, continuous documentation of tacit decisions, and alignment with a wide array of standards. The diversity of tools and staff profiles across heritage institutions and the documentation policy of each institution means that, while paradata (albeit not named as such) as a long-established practice of professionals, they often remain scattered and unsystematic as "grey literature", facing the real risk of loss when workflows are updated, or not institutionalised.

Despite this complexity, Deliverable 21 shows that many of the advances made (such as out three-tier, five-phase approach), are realistically scalable. Smaller institutions, even with limited digital capacity, can meaningfully benefit by adopting lightweight, minimum-viable tiers, modular documentation templates, and simplified process maps that do not require high-end infrastructure. The tiered approach is explicitly designed to offer an achievable entry point while providing a pathway towards more advanced practices as resources grow.

In this sense, the forthcoming WP3 final deliverable (D22: the Digitization Competence Audit Toolkit) is well positioned to provide the practical bridge that institutions need. By considering the institutional reality in the field and the day-to-day problems of professionals, WP3 aims to translate its methodology into an actionable roadmap tailored for sharing data with immersive and metaverse platforms. The final Toolkit promises to operationalize the paradata framework through helping institutions identify gaps, choose an appropriate tier, streamline decision-making, and gradually integrate paradata capture into daily digitisation work.

12. References

A) Standards, Specifications, Institutional Frameworks

Article 29 Data Protection Working Party. (2010). *Opinion 3/2010 on the principle of accountability*.

Cultural Heritage Imaging. (n.d.). *Digital Lab Notebook*.

Dublin Core Metadata Initiative. (2020). *Dublin Core metadata element set, version 1.1*.

Europeana Foundation. (n.d.-a). *Europeana Data Model (EDM) documentation*. [pro.europeana.eu]

Europeana Foundation. (2015, February 7, 2025 update). *Publishing Framework*. [pro.europeana.eu]

European Commission. (2021). *Commission Recommendation (EU) 2021/1970 of 10 November 2021 on a common European data space for cultural heritage*. *Official Journal of the European Union*. [eur-lex.europa.eu]

European Parliament & Council. (2019). *Directive (EU) 2019/790 of 17 April 2019 on copyright and related rights in the Digital Single Market*. *Official Journal of the European Union*.

European Union. (2016). *Regulation (EU) 2016/679 (General Data Protection Regulation)*. *Official Journal of the European Union*.

European Union. (2024). *Regulation (EU) 2024/1689 of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)*. *Official Journal of the European Union*. [eur-lex.europa.eu]

Federal Agencies Digital Guidelines Initiative. (2023). *Technical guidelines for digitizing cultural heritage materials (3rd ed.)*. [digitizati…elines.gov]

IIIF Consortium. (2024). *IIIF Presentation API 3.0*. [iiif.io]

International Rights Statements Working Group. (2018). *Recommendations for standardized international rights statements (v1.2.2)*. [rightsstatements.org]

International Organization for Standardization. (2012). *ISO 14721:2012 Space data and information transfer systems—Open archival information system (OAIS)—Reference model*. [iso.org]

International Organization for Standardization. (2021). *ISO 19264-1:2021 Photography—Archiving systems—Imaging systems quality analysis—Part 1: Reflective originals*. [iso.org]
Object Management

Object Management Group. (2013). Business Process Model and Notation (BPMN), Version 2.0.2. [omg.org]

PREMIS Editorial Committee. (2015). PREMIS data dictionary for preservation metadata (Version 3.0).

Van Dormolen, H. (2025). *Metamorfoze Preservation Imaging Guidelines: Image quality (Version 2.0)*. [metamorfoze.nl], [metamorfoze.nl]

B) Academic Literature

Ammirati, L., Watrous, B., Ftaita, A., Rigauts, T., & de Gruchy, M. (2025). Dive into Heritage: Paradata and metadata in an immersive digital heritage experience. In M. Ioannides, D. Baker, A. Agapiou, & P. Siegkas (Eds.), *3D Research Challenges in Cultural Heritage V: Paradata, metadata and data in digitization* (LNCS 15190). Springer. [link.springer.com]

Baca, M. (Ed.). (2016). *Introduction to metadata* (3rd ed.). Getty Publications.

Bentkowska-Kafel, A., Denard, H., & Baker, D. (Eds.). (2012). *Paradata and transparency in virtual heritage*. Ashgate.

Cassar, A., Baker, D., & Ioannides, M. (2025). From digital twin to memory twin: A holistic framework for cultural heritage documentation, interpretation, and adaptive reuse. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-9, 203–212. https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-203-2025

Dijkman, R. M., Dumas, M., & Ouyang, C. (2008). Semantics and analysis of BPMN process models. *Information and Software Technology, 50*(12), 1281–1294.

Enqvist, L. (2024). Paradata as a Tool for Legal Analysis: Utilising Data-on-Data Related Processes. In: Huvila, I., Andersson, L., Sköld, O. (eds) Perspectives on Paradata.

Knowledge Management and Organizational Learning, vol 13. Springer, Cham. https://doi.org/10.1007/978-3-031-53946-6_13

Huvila, I. (2025). Imperative of paradata. In M. Ioannides, D. Baker, A. Agapiou, & P. Siegkas (Eds.), *3D research challenges in cultural heritage V: Paradata, metadata and data in digitization* (Lecture Notes in Computer Science, Vol. 15190, pp. 1–11). Springer Nature. https://doi.org/10.1007/978-3-031-78590-0_1

Ioannides, M., Karittevli, E., Panayiotou, P., & Baker, D. (2025). Integrating paradata, metadata, and data for an effective memory twin in the field of digital cultural heritage. In M. Ioannides, D. Baker, A. Agapiou, & P. Siegkas (Eds.), *3D research challenges in cultural heritage V: Paradata, metadata and data in digitization* (Lecture Notes in Computer Science, Vol. 15190, pp. 24–35). Springer Nature. https://doi.org/10.1007/978-3-031-78590-0_3

Ioannides, M., Baker, D., Agapiou, A., Siegkas, P., & Cassar, A. (2025). Quality certification in data acquisition for cultural heritage: The power of paradata for high-quality digital 3D cultural heritage assets. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLVIII-M-9, 623–632. https://doi.org/10.5194/isprs-archives-XLVIII-M-9-2025-623-2025

Ioannides, M., Baker, D., Agapiou, A., & Siegkas, P. (Eds.). (2025). *3D research challenges in cultural heritage V: Paradata, metadata and data in digitization* (Lecture Notes in Computer Science, Vol. 15190). Springer Nature. https://doi.org/10.1007/978-3-031-78590-0

Luengo, P., Boto Varela, G., Coll Pla, S., Costa Jover, A., Samper Sosa, A., & Vivó Codina, D. (2025). Toward a trustable digitization of built heritage: The role of paradata. In M. Ioannides, D. Baker, A. Agapiou, & P. Siegkas (Eds.), *3D Research Challenges in Cultural Heritage V: Paradata, metadata and data in digitization* (LNCS 15190, pp. 139–150). Springer. [researchgate.net], [link.springer.com]

C) Projects

Bognar, M. A., Vandermeulen, B., Hameeuw, H., Toth, M. B., Christens-Barry, W. A., & Watteeuw, L. (2021). *3Pi project – KU Leuven NBMSI metadata standard (Version 2.0)*. Zenodo.

Hameeuw, H., Bognar, M. A., Vandermeulen, B., Toth, M. B., Christens-Barry, W. A., & Watteeuw, L. (2022). *3Pi project – KU Leuven NBMSI hardware document (Version 1.1.1)*. Zenodo.

Hameeuw, H., Vandermeulen, B., Toth, M. B., Christens-Barry, W. A., & Watteeuw, L. (2020). *3Pi project – KU Leuven basic sequence & nomenclature document NBMSI (Version 1.1.0)*. Zenodo.

Ioannides, M., Baker, D., Agapiou, A., & Siegkas, P. (Eds.). (2025). *3D Research Challenges in Cultural Heritage V: Paradata, metadata and data in digitization* (Lecture Notes in Computer Science, Vol. 15190). Springer. [link.springer.com]

Sketchfab. (n.d.). *Sketchfab*. [sketchfab.com]

UNOSAT. (n.d.). *UNOSAT*.

OpenHeritage3D. (n.d.). *OpenHeritage3D*.

13 Appendix

13.1 KU LEUVEN LIBRARIES A Use Case Scenario: Practical Process flow for manuscripts at KU Leuven Libraries

The exemplary use case scenario below describes the different steps of the digitization process for manuscripts, fragments and other fragile (documentary) materials. The process is modular, so that the supply can be fully tailored to the digitization demand.

1. Planning and complexity assessment

1.1 Project advice and preparation

Digitization projects are complex processes with various implementing partners, for which the implementation is preferably centrally coordinated. The intended finality of a project (answering the overarching question "what do we want to achieve through digitization?") also determines the various steps in the process, from start to completion, and must be focused in advance. Each project is therefore developed individually.

1.2 Advice, project definition and translation of finality into an action plan and necessary infrastructure

Before the start, each project step is defined and worked out in practice in order to achieve the desired result. To this end, the digitization question is analysed and an action plan is drawn up with the concrete project goal as a starting point. Extensive attention is paid to the correct translation of substantive questions into technical implementation and to the incorporation of quality checks throughout the entire process. Project development is tailor-made, but where possible, standardized processes are used.

1.3 Budget calculation and funding

Depending on the project definition and the concretization of the plan of action, a realistic price calculation is drawn up for the various process steps. This price calculation includes various aspects, such as the services used, the initial storage cost and the recurring storage cost. This involves differentiating between an internal and an external service.

1.4 Legal aspects

Digitization takes place within a legal framework. Where necessary, legal aspects are included. We draw up contracts (e.g. regarding the use of images) and take out insurance (e.g. for transport). Particular attention is paid to the definition and clarification of copyright and intellectual and property rights.

1.5 Guidance on external digitization / outsourcing

Some projects are suitable for outsourcing to a digitization company. It is important to know that only a few steps can be outsourced (e.g. conversion physical - digital; limited metadata for digitization). The importance of quality control is increasing significantly. Quotations are requested in advance, processes are drawn up, contracts are drawn up (including concrete quality conditions), and materials are physically prepared for transport and scanning.

2. Project Coordination

2.1 General Coordination

Central coordination of a digitization project increases the efficiency and quality of both the process and the end result. The coordinator coordinates the operation of the various parties involved and the individual process steps. A workflow management system is used for this purpose.

3. Metadata Strategy

3.1 Metadata advice

Both the descriptive and the structural metadata can be provided by the applicant himself or can be created by KU Leuven Libraries, Library Services. The way in which metadata is created (description depth, description model, etc.) and delivered must first be discussed with the Digitization Department. This service will advise on the modalities that are best applied in this area and with which partners further agreements may need to be made (BD and/or LIBIS).

3.2 Descriptive Metadata

Descriptive metadata is needed in every situation to interpret objects (i.e. i.e. identify) and to link the digital visual material to them (at a later stage). Digital images of a manuscript, for example, are always linked to the catalographic description (i.e. descriptive metadata) of this manuscript, so that one knows from which handwriting the images originate.

3.3 Structural Metadata

Structural metadata describes how data relates to each other and how it is “ordered”. For example, the foliation of the (digital) folios in the form of labels can be seen as structural metadata. Another example is the information about the collection in which a manuscript is housed and at which institution this collection is kept.

3.4 Preparing for and executing metadata import and/or export, and delivery of metadata to customer

After creation, the metadata can be imported and/or exported to/from a certain environment (e.g. Alma, Canto, Scope, network drive) and this in a certain format (e.g. XML, CSV). This can be done from the customer to KU Leuven Libraries or vice versa, depending on the situation. These import/export actions may need to be preceded by mapping/metadata conversion agreements between customer and supplier.

4. Conversion of physical object to digital data

A digital surrogate is created based on physical objects. This conversion is thoroughly prepared by a prior analysis of the original object and of the intended purpose of the digitization. The results of this analysis determine the way in which this conversion takes place.

4.1 Checking physical condition and any needs for preparatory conservation

The collection manager or conservator-restorer decides whether any conservation prior to digitization is necessary to preserve the integrity of the object during digitization. If conservation is necessary, the collection manager will take the necessary steps. The operator in the Digital Lab examines whether the object can be digitized safely with the existing infrastructure. Depending on the state of the binding or book block and the type of object, it is decided which infrastructure is used.

5. Data Acquisition

5.1 Standard digitization (book scanner)

Standard digitization ensures a faithful digital copy of the physical original. Specialised infrastructure is used for digitising highly fragile materials in a standardised setting. Manuscripts whose format allows it can thus be easily digitized with respect for the physical integrity of the object, thanks to a V-shaped book cradle. The images are made with cameras, which allows for better quality, faithful color reproduction and smooth processing of the images. After recording and checking, the digital images are cropped and given a unique number. Recordings can be made on location but are preferably carried out in the Digital Lab of KU Leuven Libraries. This room is specifically designed for digitization, with dark walls and ceilings, air conditioning and controlled humidity, and is secured with alarm and double lock.

5.2 High-end digitization (manual/high resolution)

High-end digitization ensures a faithful digital copy of an original at high resolution, via adapted infrastructure or with special operations (e.g. manual scanning). To this end, specialized infrastructure is used in a setting that is adapted to the object. An example is the Conservation Book Copy Stand, a construction made by Manfred Mayer (University of Graz). Manuscripts are placed in a V-shaped book cradle (110°) for safe manipulation of fragile book bindings during digitization. The position of the book cradle is adjusted to the condition of the binding, and folios are kept parallel to the camera during photography via light negative pressure. Other specialized infrastructure such as reproduction tripods or other accessories is also available to securely digitize a wide variety of materials and formats. In general, this is specialized, manual work. After recording and checking, the digital images are cropped and given a unique number. Recordings can be made on location but are preferably carried out in the Digital Lab of KU Leuven Libraries. This room is specifically designed for digitization, with dark walls and ceilings, air conditioning and controlled humidity, and is secured with alarm and double lock.

5.3 Digitization with a view to further processing

Standard digitization provides a faithful digital copy of an original that can serve as a basis for further editing in function of transcription and other specific forms of enrichment for manuscripts. Specialized infrastructure is used for this in a standardized setting. Manuscripts whose size allow it can be easily digitized while respecting the physical integrity of the object, thanks to a V-shaped cradle. Depending on the quality requirements of the application for which the images are intended, the files are specifically edited at the level of contrast and tone reproduction. After recording and checking, the digital images are cropped and given a unique number. Recordings can be made on location but are preferably carried out in the Digital Lab of KU Leuven Libraries. This room is specifically designed for digitization, with dark walls and ceilings, air conditioning and controlled humidity, and is secured with alarm and double lock.

5.4 Advanced imaging (always including high-end digitization for reference)

Advanced imaging provides additional information about the object's material features or makes elements visible that cannot be seen with the naked eye. To this end, the Digital Lab of KU Leuven Libraries works with a wide range of techniques and infrastructure that are part of VIEW, KU Leuven Core Facility for Heritage Research and Digitization Technologies (a collaboration between KU Leuven Libraries and Book Heritage Lab). The Portable Light Dome (infrastructure acquired as part of the RICH project) consists of a dome with 228 LED lamps on the inside for omni-directional lighting. The object to be monitored is located at the bottom in the middle. A high-resolution camera (28 million pixels) is attached to the dome at the top, which is aimed downwards, at the object. When shot, each LED lamp lights up separately, each time taking a photo. In this way, the object is illuminated from any angle, 360° around and with light from above to shaving light. Through processing, these 228 images are processed into an interactive file. This allows the object to be virtually re-illuminated in order to optimally visualize the topography of the object and other minute details. Calibration allows further measurements to be carried out in both the X-Y axis and the Z-axis. Export to 3D is also possible. The Portable Light Dome exists in two versions: a White Light version and a Multi Spectral version (with 5 spectral bands between UV and IR). Multispectral images are made in the context of heritage research or with a view to conservation. The images provide additional information about the material characteristics of the object (signature, pigments, retouches, etc.) or make visible elements that cannot be seen with the naked eye, such as texts that are difficult to read (censorship, palimpsest, faded text, etc.). The images are made with specific filters (14) or with LED lamps with a narrow spectral band (infrastructure acquired in the context of the 3PI project). After recording, the images are processed and edited via specific algorithms.

Transmitted light or shaving light images are made in the context of research, conservation or visualization of watermarks. They also provide additional information about the condition of the object, e.g. with regard to thickness of pigment layers, ink corrosion, etc. These images can be taken with both white and multispectral light.

Recordings can be made on location but are preferably carried out in the Digital Lab of KU Leuven Libraries. This room is specifically designed for digitization, with dark walls and ceilings, air conditioning and controlled humidity, and is secured with alarm and double lock.

5.5 Control, export and delivery of image files

After ingestion, the files are checked against the structural metadata. The control is done twice, both by the operator of the Digital Lab and by the metadator, in order to avoid errors in the process. After all, adjustments after ingesting require more resources than checking in advance. After checking, the conversion to an archive format (tif) is done for sustainable digital preservation. If ingest is part of the digitization process, files that are ready are prepared for ingest together with a checksum file and the structural metadata and then imported by LIBIS. If ingest is not included in the digitization process (e.g. for assignments for external customers), the delivery of files to the customer is done as agreed. This can be done via e-mail, Box, Dropbox, Belnet Filesender or an external drive.

6. Enrichment of digital data

Digitization makes it possible to transform physical originals into digitally usable data. In addition, the digital objects can be enriched with additional data that provide additional information, both in terms of content and its description and in terms of materiality.

6.1 OCR (Optical Character Recognition)

Through OCR or Optical Character Recognition, a digitized text becomes machine-readable. However, this technology is rarely suitable for handwriting. That is why we closely follow technological evolutions in this area, such as the development of transcription tools (e.g. Transkribus), and the possibilities of crowdsourcing. Although such tools are not currently a permanent part of our offer, we are looking for opportunities to help develop and introduce them. Do not hesitate to contact us for this.

6.2 Metadata (extension)

Enrichment of the metadata facilitates in-depth access to the digital objects. This metadata is tailored to the individual question and includes, for example, metadata at the level of the text component.

6.3 Advanced imaging

Advanced imaging provides additional information about the object's material features or makes elements visible that cannot be seen with the naked eye.

7. Storage and Preservation

Various storage environments are available for the storage and preservation of the digital objects and files. The optimal choice of storage environment depends on the specific needs and the available resources. Objects can also be exported directly to the customer for storage in the customer's storage environment.

7.1 Teneo

Teneo is a system for the sustainable preservation and accessibility of digital and digitized collections and archives.

In Teneo, the digital files are grouped into so-called "intellectual entities" (IEs). An IP is a set of files and metadata that form a logical whole with a view to sustainable preservation and accessibility, for example a digitised manuscript. There can be several representations of an IP, whereby a distinction is made between representations that are intended for long-term preservation (e.g. the original recordings), and derived representations that are made purely for the purpose of disclosure (e.g. a conversion to a lower resolution and/or a different file format). These can be generated automatically based on several parameters that can be set in advance.

7.2 Alma Cloud

Alma Cloud is an alternative module within Alma for storing and accessing digital files when long-term preservation is not required. Metadata and files (often PDF/A) are stored in a secure cloud environment, with options for manual or batch uploads linked to descriptive records for display in collections. For unique materials like manuscripts, Alma Cloud is rarely suitable for permanent storage but can serve as temporary storage, for example during restoration. KU Leuven provides guidelines for choosing between Alma Cloud and Teneo for digitization projects.

7.3 Export and delivery of data to the customer

Files can be delivered to customers in multiple ways (e.g., email, Box, Dropbox, Belnet Filesender, or hard disk). For Alma Cloud and Teneo, end users can download files or full Intellectual Entities (IEs) via standard viewers. Data suppliers may export through Rosetta UI (for KU Leuven Libraries and KADOC, after training), and LIBIS can provide tailored exports. For Canto, users can download data directly or export while preserving the database structure. Metadata export is available in CSV or XML formats, and BD Digitization offers customized exports on request.

8. Publication of Data

The digital objects can be published in one or more platforms:

8.1 Limo

Limo is a contemporary search platform for a specific database or repository, with strong integration with Alma of Scope. In addition, Limo can also be expanded into a portal that makes many information sources from various suppliers searchable simultaneously and in an intuitive way.

8.2 Viewers

The digital objects can be made accessible to the end user via different types of viewers. The viewer chosen depends on the type of material and the underlying system (Teneo or Alma Cloud). The most important viewers for objects in Teneo are the generic IE viewer, the Universal Viewer (IIIF) and the Mirador viewer (IIIF). The generic IE viewer can display all kinds of file formats (including media files) within the viewer and allows you to switch between the different representations of the object, as far as they are accessible. The IIIF viewers are mainly intended to display high-quality image files in a high-performance way. In the Mirador viewer, multiple images/objects can be displayed side by side. This viewer is only used for specific collections and projects. In addition, Teneo offers a number of special viewers for specific types of material and for collections.

8.3 Omeka/Heron

Omeka is an open-source online publishing platform with a focus on making digital collections and stories accessible to a large audience. With Omeka, digital objects are displayed in a visually attractive way with numerous possibilities such as simple tools for creating online exhibitions or informative stories, seamless integration with the collection management system and synchronization of data. In addition, a series of plug-ins are made available, for example to determine the layout or to give visitors the opportunity to give comments or additional feedback.

13.2 Heritage Malta Paradata Acquisition Sequence

ACQUISITION SESSION DETAILS

ACQUISITION SESSION INFORMATION	
Jira Number	
Description of Acquisition Session	

PEOPLE INVOLVED	
Role	**Name**
Custom Role	**Custom Name**

DATE AND LOCATION	
Date of Acquisition Session	
Acquisition Time Started	
Location of Session	

Environment Details	

METHODOLOGIES AND OTHER INFORMATION	
Methodology and Process	
Difficulties/Constraints during Acquisition	

PARADATA DETAILS

AUTOMATED AERIAL PHOTOGRAMMETRY PARADATA	
Name of Object	
Name/Number of Scan	
Image Format	
Number of Images	
Shutter Speed	
ISO	
Aperture	
Resolution	
Focal Length	

MANUAL AERIAL PHOTOGRAMMETRY PARADATA	
Name of Object	
Name/Number of Scan	
Image Format	
Number of Images	

Shutter Speed	
ISO	
Aperture	
Resolution	
Focal Length	

GROUND BASED PHOTOGRAMMETRY PARADATA	
Name of Object	
Name/Number of Scan	
Camera Setup Used	
Image Format	
Number of Images	
Shutter Speed	
ISO	
Aperture	
Resolution	
Focal Length	

SMALL ARTEFACT PHOTOGRAMMETRY PARADATA	
Name of Object	
Name/Number of Scan	
Camera Setup Used	
Image Format	
Number of Images	
Shutter Speed	
ISO	
Aperture	
Resolution	
Focal Length	

LiDAR PARADATA	
Name of Object	
Name/Number of Scan	
Number of Points	
What level of detail	

HANDHELD LASER PARADATA	
Name of Object	
Name/Number of Scan	
Resolution	

Co-funded by the European Union

No of Vertices of Raw Scan	

RTI PARADATA	
Name of Object	
Name/Number of Scan	
Image Format	
Number of Images	
Shutter Speed	
ISO	
Aperture	
Resolution	
Focal Length	

EQUIPMENT USED

DRONE EQUPMENT	
DRONE	

PHOTOGRAMMETRY EQUIPMENT Setup 1	
Camera	
Lens	
Flash	
Filter	
Setup 2	
Camera	
Lens	
Flash	
Filter	
Setup 3	
Camera	
Lens	
Flash	
Filter	

LiDAR EQUPMENT	
LiDAR	Leica RTC 360
Total Station	

HANDHELD LASER EQUIPMENT	
HandHeld Laser	Creaform Go!Scan

RTI EQUIPMENT	
Camera	
Lens	
Flash	
Filter	
RTI Setup	

13.3 Jagiellonian University Additional Processes

13.3.1 Phase 1: Planning and Complexity Assessment

Interpretation

Examples Interpretation - Phase 1 Planning & Complexity Assessment

Overview

Phase 1 sets the foundation for the entire digitization project. Poor planning here leads to problems in every subsequent phase. These 4 examples show: 2 simple (done right), 1 complex (underestimated complexity → reworks), 1 very complex (wrong tier selected → restart).

Example 1: Simple Tier 1 - Museum Vase

Summary
- **ID:** PROJ-2025-015
- **Object:** Decorative Ceramic Vase (19th century)
- **Tier:** 1 (Minimum Viable)
- **Timeline:** 1 submission, approved in 1 day
- **Complexity:** Simple

What Happened
Perfect example of a straightforward Tier 1 project. Museum knows exactly what they want: online catalog with 3D viewer.

Key Decisions:
- Clear purpose: "General Outreach - online museum catalog"
- Appropriate budget: €500 (realistic for single object, basic photogrammetry)
- Simple object: Ceramic vase, controlled environment, no access restrictions

Why It Worked:
- Scope matched tier (Tier 1 = basic documentation, that's all they need)
- Purpose singular and clear (not trying to do everything)
- Budget realistic
- Object straightforward (no complexity challenges)

Approved: 2025-01-09 (1 day after submission)

Lesson: When you know what you want and keep it simple, Phase 1 is fast.

Example 2: Simple Tier 2 - Renaissance Sculpture

Summary
- **ID:** PROJ-2024-089
- **Object:** Renaissance Sculpture - Virgin Mary (16th century)
- **Tier:** 2 (Scientific)
- **Timeline:** 1 submission, approved in 2 days
- **Complexity:** Simple

What Happened
Well-planned Tier 2 for conservation and research.

Key Decisions:
- Dual purpose clearly stated: Conservation planning + Research
- Target quality specified: Europeana Tier 4 (research-grade)
- Output formats match uses: High-res archive (5M polygons) + web viewer (glTF)
- Budget appropriate: €15,000 from Ministry of Culture

Tier 2 Additions:
- Target quality quantified (not just "high quality")
- Multiple output formats for different audiences
- Purpose drives quality requirements

Why It Worked:
- Purpose justified the tier (research = Tier 2 minimum)
- Budget matched ambition

- Team experienced (National Heritage Institute)

Approved: 2024-12-22 (2 days after submission)

Lesson: Tier 2 needs more detail than Tier 1, but with clear purpose and adequate budget, still straightforward.

Example 3: Complex Tier 3 - Castle Ruins (3 Attempts)

Summary
- **ID:** PROJ-2025-007
- **Object:** Medieval Castle Ruins (13th century, 45m × 30m site)
- **Tier:** 3 (Memory Twin)
- **Timeline:** 3 attempts over 8 days
- **Complexity:** Complex - Initially underestimated

Timeline
- **Jan 5:** Attempt 1 → ❌ Needs Rework (superficial complexity assessment)
- **Jan 9:** Attempt 2 → ❌ Needs Rework (better but missing critical details)
- **Jan 13:** Attempt 3 → ✅ Approved (comprehensive documentation)

Attempt 1 - Too Superficial

What Was Submitted:
- Material complexity: "Stone (various weathering levels)" - vague
- Geometric complexity: "Complex - many cavities" - not quantified
- Environment: "Outdoor site with vegetation" - no detail

Rejection:
> "Complexity assessment too superficial for Tier 3. Need to specify moss/lichen coverage. 'Many cavities' not quantified - need inaccessible percentage. Vegetation vague - specify removal plan, seasonal timing, access restrictions."

Problems:
- Tier 3 requires **quantified** assessments, not vague descriptions
- Missing practical planning: vegetation removal, lighting challenges, weather windows
- No mention of safety, regulations, or logistics

Team's Mistake: Thought describing complexity was enough. Tier 3 needs **actionable planning**.

Attempt 2 - Better But Still Incomplete

What Changed:
- ✅ Quantified: "~40% moss/lichen coverage", "70% accessible, 30% inaccessible"
- ✅ Detailed capture strategy: Drone + ground + scaffolding
- ✅ Weather window identified: April-May
- ✅ Vegetation removal contractor arranged

Still Missing:
- ❌ Color calibration strategy (variable lighting on stone)
- ❌ Drone regulations documentation
- ❌ Contingency plan if weather extends beyond May
- ❌ **Critical:** Safety assessment (scaffolding in ruins = structural engineer needed)
- ❌ Budget breakdown

Rejection:
> "Much improved but missing critical Tier 3 details. Need drone regulations documentation. No contingency for weather delays. CRITICAL: No safety assessment - scaffolding in ruins requires structural engineer sign-off."

Team's Realization: Tier 3 isn't just about capture quality - it's about **risk mitigation** and **regulatory compliance**.

Attempt 3 - Comprehensive (Approved!)

What Changed:
- ✅ **Safety:** Structural engineer assessment (Dr. Eng. Piotr Zieliński, dated 2024-12-15)
- ✅ **Regulations:** Drone permit from CAA (#DRONE-2025-IAE-007), airspace coordination
- ✅ **Color calibration:** X-Rite ColorChecker at 5m intervals
- ✅ **Contingency:** If April unfavorable, extend to May 15 (before tourism peak)
- ✅ **Budget breakdown:** €85K across phases with €12K contingency
- ✅ **Insurance:** €2M liability, €50K equipment
- ✅ **Quantified everything:** GSD targets (1mm close-range, 5mm aerial), accuracy goals (±2mm/±5cm)

Approval:

> "Tier 3 planning now comprehensive and realistic. All complexity dimensions thoroughly assessed with quantified metrics. Safety documentation complete. Regulatory compliance addressed. This level of planning significantly reduces Phase 3 capture risks. Exemplary Tier 3 planning."

Result: ✅ Approved

Key Lessons from Example 3

What Went Wrong:
- Underestimated what "Tier 3 planning" means
- Initially thought description = planning (wrong - need actionable plans)
- Missed that complexity assessment must address **risks** not just challenges

What Went Right:
- Team listened to feedback and added detail iteratively
- Each attempt significantly improved
- Final version is actually **usable** in Phase 3 (contractor can bid from this)

Cost of Rework:
- Time: +8 days (3 iterations)
- BUT: Saved weeks in Phase 3 (no mid-capture surprises)

Critical Insight:
> **Tier 3 planning is risk management.** You're not just describing the object - you're identifying every challenge and documenting how you'll address it. Safety, regulations, weather, access, budget - all must have concrete plans.

Example 4: Very Complex - Wrong Tier (Phase 1 Return)

Summary
- **ID:** PROJ-2024-178
- **Object:** Historic City Center - 15 buildings
- **Initial Tier:** 3 (wrong!)
- **Final Tier:** 1 (revised)
- **Timeline:** Return to Phase 1 start + workshop + resubmission (31 days total)
- **Complexity:** Overly ambitious → rescoped

Timeline
- **Nov 15:** Attempt 1 (Tier 3) → ❌ **Insufficient Plan** (return to Phase 1 start)

- **Nov 18:** Decision to return to Phase 1
- **Dec 5:** Team workshop (stakeholder alignment)
- **Dec 15:** Attempt 2 (Tier 1 revised) → ✅ Approved

Attempt 1 - Fundamentally Flawed

What Was Proposed:
- 15 buildings, interiors + exteriors
- Tier 3 (Memory Twin)
- Budget: €25,000
- Timeline: 3 months
- Multiple purposes: Urban planning + heritage + tourism + research
- Team: Urban planner (no digital heritage experience)

Critical Decision: Return to Phase 1 START

Reviewer's Analysis:
> "FUNDAMENTAL ISSUES - Cannot proceed: (1) WRONG TIER: Tier 3 'Memory Twin' is for SINGLE objects. This is 15 BUILDINGS - Tier 3 would require documenting EVERY material challenge in EVERY room. Not feasible. (2) UNREALISTIC BUDGET: €25K cannot cover Tier 3 for 15 buildings. Estimate: €200-300K minimum. (3) SCOPE CONFUSION: Multiple contradictory purposes. (4) TIMELINE IMPOSSIBLE: 2 days per building including interiors. (5) TEAM INADEQUATE: No digital heritage expertise."

Why Return to Phase 1 START (not just rework):
- Tier selection wrong (fundamental error)
- Budget off by 10x
- Timeline impossible
- Team lacks expertise
- Multiple incompatible purposes

This isn't "fix some details" - this is "restart the entire planning process."

The Workshop (Dec 5)

Participants:
- City Planning Office (requester)
- National Heritage Institute consultant (expert)
- Budget Office (reality check)

Outcomes:
1. **Primary purpose identified:** Urban planning visualization (facades only)
2. **Tier revised:** Tier 1 (exteriors only, no interiors)
3. **Budget reassessed:** €28,000 (€25K + €3K consultant)
4. **Timeline:** 4 months (realistic)
5. **Scope redefined:** Street facades + courtyard facades (NO interiors)
6. **Consultant engaged:** Dr. Anna Lewandowska

Key Insight: Team realized they wanted **urban planning tool**, not **heritage research archive**. Different purposes = different tiers.

Attempt 2 - Rescoped to Tier 1 (Approved!)

What Changed:
- ✅ Tier: 1 (facades only)
- ✅ Purpose: Primary = urban planning, secondary = heritage baseline
- ✅ Scope: EXTERIOR ONLY (no interiors)
- ✅ Budget: €28K (realistic for facade capture)
- ✅ Timeline: 4 months (achievable)
- ✅ Team: Added consultant for guidance

Approval:
> "Excellent course correction. Team showed maturity by accepting feedback and rescoping appropriately. Tier 1 (facades only) is realistic. Purpose now clear and focused. This demonstrates proper use of Phase 1 review - better to restart planning than commit to unachievable scope."

Result: ✅ Approved

Key Lessons from Example 4

What Went Wrong:
- **Tier confusion:** Thought "big/important project = Tier 3" (WRONG - tier is about documentation depth, not importance)
- **Scope creep:** Wanted everything (planning + heritage + tourism + research)
- **Budget disconnect:** €25K vs €200K+ reality
- **Wrong expertise:** Urban planner leading digital heritage project

What Went Right:
- **Reviewer caught it early** (before money wasted in Phase 3)

- **Team accepted feedback** (no defensive response)
- **Workshop productive** (aligned stakeholders)
- **Rescoped appropriately** (facades = useful for their actual need)

Cost Analysis:
If Continued	If Restarted
€25K wasted on impossible project	+€3K consultant fee
Failure after Phase 3	Success with Tier 1
No usable output	Usable facades for planning
Team demoralized	Team educated

Critical Insight:
> **Tier selection is about PURPOSE and DEPTH, not object importance.** A major project (15 buildings) can be Tier 1 if purpose is visualization. A single object can be Tier 3 if research requires full complexity documentation. Match tier to purpose, not prestige.

Comparison Table

Aspect	Example 1	Example 2	Example 3	Example 4
Object	Ceramic Vase	Sculpture	Castle Ruins (site)	15 Buildings
Tier	1	2	3	3 → 1 (revised)
Complexity	Simple	Simple	Complex	Overly ambitious
Submissions	1	1	3	2 + workshop
Days	1	2	8	31
Budget	€500	€15,000	€85,000	€25K → €28K
Main Issue	None	None	Underestimated Tier 3 detail	Wrong tier
Resolution	Immediate approval	Immediate approval	Iterative improvement	Restart with new tier
Key Lesson	Keep it simple	Match tier to purpose	Tier 3 = risk management	Tier ≠ importance

---

General Patterns

Success Factors (Examples 1-2)
1. **Clear singular purpose**
2. **Appropriate tier selection**
3. **Realistic budget for chosen tier**
4. **Experienced team**

Common Tier 3 Mistakes (Example 3)
1. **Vague assessments** ("complex", "challenging" without quantification)
2. **Missing safety/regulatory documentation**
3. **No contingency planning**
4. **Underestimating documentation depth**

When to Return to Phase 1 Start (Example 4)
- Wrong tier selected
- Budget off by order of magnitude
- Multiple incompatible purposes
- Team lacks necessary expertise
- Fundamental scope confusion

Time Estimates by Tier

- **Tier 1:** 1-2 days (if straightforward)
- **Tier 2:** 2-4 days (needs quality specs)
- **Tier 3 (first attempt):** 5-10 days (detailed assessment needed)
- **Tier 3 (with reworks):** Add 2-3 days per iteration
- **Phase 1 restart:** 20-30 days (includes stakeholder workshop)

Recommendations

For Tier 1:
- Be specific about purpose (not just "general outreach")
- Don't over-plan - Tier 1 is intentionally minimal

For Tier 2:
- Quantify target quality (Europeana tier? accuracy level?)
- Specify output formats for different uses
- Justify why Tier 2 (what research/conservation need?)

For Tier 3:
- **Quantify everything:** percentages, measurements, counts
- **Address all risks:** safety, regulations, weather, access
- **Include documentation:** permits, engineer sign-offs, insurance
- **Budget breakdown:** show it's realistic
- **Contingency plans:** what if timeline slips?

For Tier Selection:
- **Tier = documentation depth, not object importance**
- Match tier to primary purpose:
  - Tier 1: Visualization, outreach, general access
  - Tier 2: Research, conservation, scholarly publication
  - Tier 3: Full provenance, complex research, reproducibility critical
- If purpose unclear → workshop with stakeholders first
- If budget << tier requirement → reduce tier or increase budget

---

Document Version: 1.0
Date: 2026-01-16
Purpose: Evaluation package for Phase 1 BPMN process

Quick Guide
Phase 1: Planning & Complexity Assessment - Quick Guide

🎯 For Evaluators

This package contains:
- **BPMN Diagram:** `01_Phase1_Planning_ComplexityAssessment_1.bpmn` - Process flow with 3-tier structure
- **Example Data:** `phase1_examples_data.xml` - 4 real-world scenarios (2 simple, 2 complex)
- **Interpretations:** `phase1_interpretation.md` - Detailed analysis including wrong tier selection
- **Validation Checklist:** `phase1_validation_checklist.md` - Review criteria
- **This Guide:** Quick reference for process usage

Quick Start: Open BPMN diagram, then read Example 4 interpretation to understand tier selection issues.

📋 Overview

Purpose: Plan digitization project and assess object complexity before any capture begins
Duration: 1-8 days depending on tier
Output: Planning Paradata Record (XML)

This is the MOST IMPORTANT phase - poor planning here causes problems in every subsequent phase.

🎯 What This Phase Does

Phase 1 answers fundamental questions:
- What are we digitizing?
- Why? (purpose determines tier and quality)
- What tier of documentation do we need?
- How complex is this going to be? (Tier 3 only)
- Is this feasible with our budget and timeline?

Key Decision: Tier selection (1, 2, or 3) based on purpose, not object importance.

🏊 Three-Tier Structure

Tier 1: Minimum Viable
Who needs this: General public access, visualization, outreach
Tasks:
1. Record Institution (name, address, contact)
2. Record Object Type (be specific)
3. Record Object Size (measurements)
4. Assign Project ID
5. Define Project Scope (budget, timeline, deliverables)
6. Define Digitization Purpose (specific use case)

Example:
- Object: Museum vase
- Purpose: Online catalog 3D viewer
- Budget: €500
- Tier 1 sufficient ✅

Tier 2: Scientific
Who needs this: Research, conservation, scholarly publication
Additional Tasks:
7. Set Target Quality (Europeana tier? Accuracy level?)
8. Define Intended Output Format (formats for different uses)

Why Tier 2: Need research-grade quality, reproducibility, or archival longevity.

Example:
- Object: Renaissance sculpture
- Purpose: Conservation planning + research publication
- Target: Europeana Tier 4, ±2mm accuracy
- Formats: High-res archive (OBJ) + web viewer (glTF)
- Tier 2 required ✅

Tier 3: Memory Twin
Who needs this: Complex research requiring full provenance, high-risk captures, reproducibility-critical projects
Additional Tasks:
9. Assess Material Complexity (optical challenges, preparation needs)
10. Assess Geometric Complexity (accessibility, occlusion strategy)
11. Assess Environment Complexity (access, weather, safety, regulations)

Why Tier 3:
- Object materially/geometrically complex (transparency, deep cavities, large scale)
- Site-based (outdoor, access restrictions, weather-dependent)
- High-risk (safety concerns, expensive equipment)
- Full provenance required (every decision documented)

Example:
- Object: Medieval castle ruins (45m × 30m site)
- Purpose: VR reconstruction + research
- Challenges: 30% inaccessible, outdoor, public site, weathering
- Requires: Safety assessment, permits, weather contingency
- Tier 3 required ✅

🚀 How to Use

Step 1: Select Tier (Based on Purpose)
Tier is about documentation depth, not object importance!

Purpose	Recommended Tier
Online catalog, web viewer, general outreach	Tier 1
Research publication, conservation, archival	Tier 2
Complex research, full provenance, high-risk site	Tier 3

Step 2: Complete Required Tasks

- **Tier 1:** Complete 6 basic tasks
- **Tier 2:** Complete Tier 1 + 2 quality tasks
- **Tier 3:** Complete Tier 1 + 2 + 3 complexity assessments

Step 3: Quality Check
For Tier 3 especially:
- Are complexity assessments **quantified**? (percentages, not "complex")
- Is safety addressed? (engineer sign-off if needed)
- Are permits documented? (drone, site access)
- Is budget realistic? (with contingency)
- Is there a contingency plan? (weather, delays)

Step 4: Approval
Reviewer decides:
- ✅ **Approved** → Proceed to Phase 2
- ⚠️ **Needs Rework** → Fix specific issues, resubmit
- 🔙 **Return to Phase 1 Start** → Fundamental issues (wrong tier, unrealistic scope)

📊 Decision Points

Gateway 1: "Tier ≥ 2?"
- **If Tier 1 only** → Skip to Approval
- **If Tier 2 or 3** → Continue to quality specification

Gateway 2: "Tier = 3?"
- **If Tier 2 only** → Skip to Approval
- **If Tier 3** → Continue to complexity assessment

💾 Output Format

File: `planning_paradata_[ProjectID].xml`
Schema: Paradata Planning Schema v1.0
Encoding: UTF-8

Example filename: `planning_paradata_IAE-2025-CASTLE-007.xml`

📝 Example: Tier 3 Record

```xml
<PlanningData phase="1">
 <Project>
  <ProjectID>IAE-2025-CASTLE-007</ProjectID>
  <ObjectName>Medieval Castle Ruins</ObjectName>
  <SelectedTier>3</SelectedTier>
 </Project>

 <Tier1Data>
  <Institution>Institute of Archaeology, Poznań</Institution>
  <ObjectType>Archaeological site - castle ruins</ObjectType>
  <ObjectSize>45m × 30m footprint, 4-12m height</ObjectSize>
  <ProjectScope>VR + research, €85K, 6 months</ProjectScope>
  <DigitizationPurpose>VR reconstruction + research publication</DigitizationPurpose>
 </Tier1Data>

 <Tier2Data>
  <TargetQuality>Europeana Tier 4, ±2mm accuracy</TargetQuality>
  <IntendedOutputFormat>Point cloud (E57), mesh (OBJ), VR (FBX)</IntendedOutputFormat>
 </Tier2Data>

 <Tier3Data>
  <MaterialComplexity>
   <Assessment>Challenging (4/5)</Assessment>
   <BiologicalGrowth>40% moss/lichen coverage</BiologicalGrowth>
   <ColorCalibration>X-Rite ColorChecker at 5m intervals</ColorCalibration>
  </MaterialComplexity>

  <GeometricComplexity>
   <Assessment>Very Complex (5/5)</Assessment>
   <Accessibility>70% accessible, 30% inaccessible</Accessibility>
   <Strategy>Drone + ground + scaffolding</Strategy>
  </GeometricComplexity>

  <EnvironmentComplexity>
   <Assessment>High (4/5)</Assessment>
   <WeatherWindow>April 1-30, contingency to May 15</WeatherWindow>
   <Safety>Structural engineer sign-off obtained</Safety>
   <Permits>Drone permit CAA #DRONE-2025-IAE-007</Permits>
  </EnvironmentComplexity>
 </Tier3Data>
```

```
  <Approval>
    <Status>Approved</Status>
    <Date>2025-01-14</Date>
  </Approval>
</PlanningData>
```

⚠️ Common Mistakes

Tier Selection Errors
❌ **Don't:** Select Tier 3 because object is important/expensive
✅ **Do:** Select tier based on documentation depth needed for purpose

❌ **Don't:** "We want best quality so Tier 3"
✅ **Do:** Match tier to actual use (web display = Tier 1 fine)

Tier 3 Assessment Errors
❌ **Don't:** "Complex geometry with many cavities"
✅ **Do:** "70% accessible, 20% needs scaffolding, 10% inaccessible"

❌ **Don't:** "Outdoor site with vegetation"
✅ **Do:** "Public park, closed March-April, vegetation removal Feb 2025, capture window April with May contingency"

Budget Realism
❌ **Don't:** €25K for Tier 3 multi-building project
✅ **Do:** Budget reality check: Tier 3 single building = €20K+, multi-building = reconsider tier

Scope Clarity
❌ **Don't:** "Planning + heritage + tourism + research + conservation"
✅ **Do:** "Primary: Urban planning. Secondary: Heritage baseline."

🔗 Related Phases

- **Next:** Phase 2 (Metadata Strategy)
- **May return from:** Phase 2, 3, 4, 5 (if fundamental planning issues discovered)
- **Critical:** Get this right - every other phase depends on it

---

📞 Standards Referenced

- London Charter: Purpose determines methodology
- VIGIE 2020/654: Complexity assessment framework
- Europeana Publishing Framework: Quality tiers
- Seville Principles: Planning before intervention

Version: 1.0
Last Updated: 2026-01-16

📊 Example Scenarios Comparison

Aspect	Example 1	Example 2	Example 3	Example 4
Object	Ceramic Vase	Renaissance Sculpture	Castle Ruins (site)	15 Buildings
Tier	1 - Minimum	2 - Scientific	3 - Memory Twin	3 → 1 (revised)
Complexity	⭐ Simple	⭐⭐ Simple	⭐⭐⭐ Complex	⭐⭐⭐⭐ Overly ambitious
Attempts	1	1	3	2 + workshop
Total Time	1 day	2 days	8 days	31 days
Budget	€500	€15,000	€85,000	€25K → €28K
Main Challenge	None	None	Underestimated Tier 3 detail	Wrong tier selection
Rework Reason	-	-	Vague complexity assessment (2x)	Tier 3 inappropriate for scale
Final Status	✅ Approved	✅ Approved	✅ Approved	✅ Approved (as Tier 1)
Key Lesson	Keep simple projects simple	Match tier to purpose	Tier 3 = quantified risk management	Tier ≠ importance

Scenarios Explained:

Example 1 & 2 (Simple - Happy Path):
- Clear purpose, appropriate tier
- Budget matches scope
- Approved immediately
- **Use these as templates**

Example 3 (Complex - Underestimated):
- Initially vague Tier 3 complexity assessment
- 2 reworks to add detail: quantification, safety, permits
- Shows what "Tier 3 planning" really means
- **Use this to understand** Tier 3 documentation depth

Example 4 (Very Complex - Wrong Tier):
- Started as Tier 3 for 15 buildings (inappropriate)
- Budget 10x too low
- Returned to Phase 1 START for tier revision
- Rescoped as Tier 1 facades (realistic)
- **Use this to understand** when to return to Phase 1 start

For detailed analysis, see `phase1_interpretation.md`

Validation Checklist

Validation Checklist - Phase 1 Planning & Complexity Assessment

Purpose
Validate that Phase 1 planning is complete, realistic, and appropriate for the selected tier before proceeding to Phase 2.

TIER 1: Minimum Viable (Required for ALL projects)

Basic Information
- [ ] Institution name, address, contact information provided
- [ ] Object type clearly described (not just "artifact")
- [ ] Object size quantified (measurements, not "small/large")
- [ ] Project ID assigned (unique identifier)
- [ ] Project scope defined (deliverables, timeline, funding source)
- [ ] Digitization purpose stated (specific use case, not generic)

Scope & Purpose Checks
- [ ] Purpose is **singular and clear** (not "outreach + research + tourism + conservation")
- [ ] Scope matches budget (realistic expectations)
- [ ] Timeline reasonable for scope

Red Flags:
- ⚠️ Vague purpose ("general use", "various purposes")

- ⚠️ Budget << scope (€500 for complex multi-building project)
- ⚠️ Multiple incompatible purposes
- ⚠️ "TBD" in any field

Decision: Tier 1 complete? **YES / NO / NEEDS REWORK**

TIER 2: Scientific (If project tier ≥ 2)

Target Quality
- [ ] Quality target specified (e.g., "Europeana Tier 4", "±2mm accuracy")
- [ ] Quality justified by purpose (why this level needed?)
- [ ] Quality is **quantified** (not just "high quality")

Intended Output Format
- [ ] Specific formats listed (OBJ, PLY, glTF, etc.)
- [ ] Formats match intended uses (archive vs web vs VR)
- [ ] Multiple formats if multiple audiences

Purpose Justification
- [ ] Clear why Tier 2 needed (research? conservation? publication?)
- [ ] Tier matches purpose (visualization = Tier 1 usually enough)

Warning Signs:
- ⚠️ "High quality" without metrics
- ⚠️ Single format when multiple uses mentioned
- ⚠️ Tier 2 selected but purpose is general outreach (consider Tier 1)

Decision: Tier 2 complete? **YES / NO / NEEDS REWORK**

TIER 3: Memory Twin (If project tier = 3)

Material Complexity Assessment
- [ ] Assessment is **quantified** (percentages, measurements, counts)
- [ ] Optical challenges identified (reflectivity, transparency, biological growth)
- [ ] Preparation needs specified (moss removal, cleaning, access prep)
- [ ] Color calibration strategy defined (if variable lighting)

Examples of good vs bad:
- ❌ "Stone with moss"
- ✅ "Limestone, 40% surface moss/lichen coverage, removal planned Feb 2025"

Geometric Complexity Assessment
- [ ] Accessibility quantified ("70% accessible, 30% inaccessible")
- [ ] Structural challenges detailed (cavities, undercuts, moving parts)
- [ ] Occlusion strategy described (drone, scaffolding, multiple angles)
- [ ] Capture strategy addresses inaccessible areas

Examples:
- ❌ "Complex geometry with cavities"
- ✅ "70% accessible ground-level, 20% needs scaffolding, 10% structurally unsafe"

Environment Complexity Assessment
- [ ] Site access restrictions documented (public site? permits needed?)
- [ ] Environmental constraints identified (weather, lighting, vegetation)
- [ ] Seasonal timing specified (optimal capture window)
- [ ] Contingency plan if timeline extends

Examples:
- ❌ "Outdoor site with vegetation"
- ✅ "Public park, closed March-April, vegetation removal Feb, capture April, contingency to May 15"

Safety & Regulatory (CRITICAL for Tier 3)
- [ ] Safety assessment completed (if elevated work, structural concerns)
- [ ] Regulatory permits obtained or in progress (drone, site access, etc.)
- [ ] Insurance coverage specified (liability, equipment)
- [ ] First aid / emergency plan (if remote or hazardous site)

Budget & Resources
- [ ] Budget breakdown by phase (not just total)
- [ ] Contingency fund included (10-15% recommended)
- [ ] Team expertise documented (who will do what?)
- [ ] Equipment needs identified

Red Flags for Tier 3:
- 🚩 Vague assessments ("complex", "challenging" without details)
- 🚩 No quantification (percentages, measurements)
- 🚩 Missing safety documentation (scaffolding = engineer needed!)
- 🚩 No regulatory permits mentioned (drone near airport? site permit?)
- 🚩 No contingency plan (weather delays? access problems?)
- 🚩 Budget unrealistic for Tier 3 scope

Decision: Tier 3 complete? **YES / NO / NEEDS REWORK**

CROSS-TIER CHECKS

Tier Selection Appropriateness
- [ ] Tier matches primary purpose
  - Visualization/outreach → Usually Tier 1
  - Research/conservation → Usually Tier 2+
  - Complex research/full provenance → Tier 3
- [ ] Budget realistic for selected tier
  - Tier 1: €500-5K (single simple object)
  - Tier 2: €5K-50K (depending on size/complexity)
  - Tier 3: €20K-200K+ (high complexity, full documentation)
- [ ] Team has expertise for selected tier
  - Tier 3 requires digital heritage expertise (not just photography)

Scope Realism Check
- [ ] Scope achievable with stated budget
- [ ] Timeline realistic (not 2 days per building for interiors)
- [ ] Team adequate for scope (multi-building = multi-person team)
- [ ] Single purpose OR clearly prioritized purposes

Consistency
- [ ] Object type matches size description
- [ ] Purpose justifies tier selection
- [ ] Budget aligns with timeline and scope
- [ ] No contradictions in documentation

APPROVAL DECISION MATRIX

Situation	Decision	Action
All tier requirements met, scope realistic	✅ **Approved**	Proceed to Phase 2
Missing/incomplete fields, fixable	⚠️ **Needs Rework**	Specific feedback, return to Phase 1
Wrong tier for purpose	🔙 **Return to Phase 1 Start**	Tier selection workshop
Budget unrealistic by order of magnitude	🔙 **Return to Phase 1 Start**	Budget reality check, rescope
Multiple incompatible purposes	🔙 **Return to Phase 1 Start**	Stakeholder alignment workshop
Team lacks necessary expertise	🔙 **Return to Phase 1 Start**	Engage consultant or revise tier

---

WHEN TO RETURN TO PHASE 1 START

Use "Return to Phase 1 Start" (not just "Needs Rework") when:

1. **Wrong Tier Selected**
  - Tier 3 for visualization purpose → Should be Tier 1
  - Multi-building project as Tier 3 → Tier not appropriate for scale

2. **Budget Disconnect**
  - Budget off by 5x or more (€25K for €200K project)
  - No budget breakdown for Tier 3

3. **Scope Confusion**
  - Multiple incompatible purposes (planning + heritage + tourism + research)
  - No clear primary purpose

4. **Team Inadequacy**
  - Tier 3 project led by team with no digital heritage experience
  - No consultant or expert on team for complex project

5. **Fundamental Misunderstanding**
  - Tier selection based on object importance, not documentation depth
  - Timeline impossible (2 days per building interiors)

Returning to Phase 1 Start requires:
- Stakeholder workshop to align expectations
- Budget reality check
- Purpose clarification
- Possible tier revision
- Expert consultation if needed

FEEDBACK BEST PRACTICES

For "Needs Rework" (fixable issues)
✅ **Do:**
- List specific missing quantifications
- Explain why detail is needed ("for contractor bidding", "for risk assessment")
- Provide examples: "Instead of 'complex geometry', specify '70% accessible, 30% inaccessible'"

- Estimate time to fix

❌ **Don't:**
- Say "insufficient detail" without examples
- Reject for minor omissions if 95% complete

For "Return to Phase 1 Start" (fundamental issues)
✅ **Do:**
- Explain why this is fundamental (not just missing detail)
- Suggest workshop or consultation
- Provide tier selection guidance
- Be constructive ("Consider Tier 1 for facades only - realistic for budget")

❌ **Don't:**
- Use harsh language ("this is ridiculous")
- Reject without guidance on how to restart
- Blame team (learning process)

EXAMPLES OF GOOD FEEDBACK

Needs Rework (Tier 3)
> "NEEDS REWORK: Complexity assessment needs quantification. Material: Specify moss coverage percentage and removal plan with contractor/date. Geometric: Replace 'many cavities' with accessibility percentage (e.g., '70% accessible'). Environment: Specify capture window with contingency plan. Missing: Safety documentation if scaffolding used (structural engineer sign-off required). Estimated fix: 2-3 days."

Return to Phase 1 Start
> "RETURN TO PHASE 1 START: Tier 3 selected but not appropriate for this project. Tier 3 'Memory Twin' is for single objects with full complexity documentation - this is 15 buildings. Recommend stakeholder workshop to: (1) Clarify primary purpose, (2) Consider Tier 1 for facades only (realistic with €28K budget), (3) Engage digital heritage consultant. Current scope (15 building interiors at Tier 3, €25K, 3 months) not feasible - would require €200K+ and 12+ months."

Approved
> "APPROVED: Phase 1 planning complete and realistic. Tier 3 complexity assessment comprehensive with quantified metrics (40% moss coverage, 70% accessible, etc.). Safety documentation excellent (structural engineer sign-off, insurance, permits). Budget breakdown shows €12K contingency. Capture window with contingency plan. This level of detail significantly reduces Phase 3 risks. Ready for Phase 2: Metadata Strategy."

TIME ESTIMATES FOR REVIEW

- **Tier 1:** 5-10 minutes (basic checks)
- **Tier 2:** 10-15 minutes (quality specs check)
- **Tier 3:** 30-45 minutes (complexity assessment detailed review)
- **Re-review after rework:** 10-15 minutes (focus on changes)
- **Phase 1 restart review:** 20-30 minutes (ensure fundamental issues resolved)

COMMON PITFALLS

1. **Tier confusion:** Assuming big project = Tier 3 (wrong - tier = documentation depth)
2. **Vague complexity:** "Complex geometry" without quantification
3. **Missing safety:** Scaffolding without engineer sign-off
4. **No contingency:** Tight timeline with no buffer
5. **Budget optimism:** Tier 3 ambition with Tier 1 budget
6. **Purpose multiplicity:** Trying to achieve 4 different goals

---

FINAL CHECKLIST

- [ ] I have reviewed all required fields for the project's tier
- [ ] Tier selection is appropriate for stated purpose
- [ ] Budget realistic for scope and tier
- [ ] Team has necessary expertise (or consultant engaged)
- [ ] If Tier 3: complexity quantified, safety addressed, permits documented
- [ ] Scope is achievable (not overly ambitious)
- [ ] My feedback is specific and constructive
- [ ] Decision (Approved/Needs Rework/Return to Start) clearly stated

Reviewer: _______________
Date: _______________
Decision: Approved / Needs Rework / Return to Phase 1 Start

Checklist Version: 1.0
Last Updated: 2026-01-16
Purpose: Standardize Phase 1 validation

13.3.2 Phase 2: Metadata Strategy

Interpretation

Examples Interpretation - Phase 2 Metadata Strategy

Overview

Phase 2 defines HOW you'll describe your digital objects. Metadata strategy decisions affect discoverability, interoperability, and long-term usability. Poor metadata planning creates technical debt for decades.

Example 1: Simple Tier 1 - Museum Vase (Dublin Core)

Summary: PROJ-2025-015, Ceramic Vase, Tier 1, 1 day approval

Decision: Dublin Core Simple - sufficient for online catalog.

Why It Worked:
- Purpose: Basic online catalog (Tier 1 = basic metadata)
- Standard appropriate: DC adequate for simple discovery
- Language: Polish (institutional language)
- No complexities

Approved in 1 day. Simple projects stay simple.

Example 2: Simple Tier 2 - Sculpture (LIDO + Vocabularies)

Summary: PROJ-2024-089, Renaissance Sculpture, Tier 2, 1 day approval

Key Decisions:
- **LIDO v1.1** (not Dublin Core) - research-grade standard
- **Bilingual:** English primary (international research), Polish secondary
- **Controlled vocabularies:** Getty AAT, GeoNames, PeriodO
- **Versioning:** Documented migration from 1987 card catalog
- **Roles:** Curator (content) + Metadata specialist (encoding)

Why LIDO not DC: Research purpose requires structured, rich metadata. LIDO designed for museum objects, supports complex relationships.

Approved in 1 day. Tier 2 needs more detail but straightforward with clear standards.

Example 3: Complex Tier 3 - Castle (Multilingual + EDM, 3 Attempts)

Summary: PROJ-2025-007, Castle Ruins, Tier 3, 3 attempts over 8 days

Timeline
- **Jan 15:** Attempt 1 → ❌ Vague standards, incomplete
- **Jan 19:** Attempt 2 → ❌ Better but missing critical details
- **Jan 23:** Attempt 3 → ✅ Comprehensive strategy

Attempt 1 - Too Vague

Problems:
- "LIDO + Dublin Core" - which is primary? Unclear relationship.
- "English and Polish" - translation workflow undefined
- "Local terms" - not documented
- EDM mentioned - no mapping strategy

Rejection: Everything mentioned, nothing detailed.

Attempt 2 - Better But Incomplete

Improvements:
- ✅ LIDO primary, DC derived
- ✅ Polish primary, English full, German summary
- ✅ Local vocabulary named with URI
- ✅ LIDO→EDM mapping mentioned

Still Missing:
- ❌ Translation workflow (who? when? QC?)
- ❌ EDM crosswalk table (too vague)
- ❌ Update policy (errors post-publication?)
- ❌ Rights strategy for metadata itself

Attempt 3 - Comprehensive (Approved!)

Complete Tier 3 Strategy:

1. Standards Hierarchy:
- LIDO v1.1 (primary/authoritative)
- Dublin Core (auto-derived for web)
- Clear relationship documented

2. Multilingual Workflow:
- Sequential process with timeline
- Quality control by content authority
- 7-day workflow documented

3. Controlled Vocabularies:
- All with persistent URIs (Getty AAT IDs, GeoNames IDs)
- Local vocabulary with institutional URI
- Machine-actionable (not just human-readable names)

4. EDM Aggregation:
- Complete crosswalk table documented
- Link to detailed mapping PDF
- Each LIDO element → EDM property specified

5. Metadata Rights:
- Metadata itself: CC0 (public domain)
- Descriptive text: CC BY 4.0
- Digital objects: CC BY-NC-SA 4.0
- Complies with RightsStatements.org

6. Update Policy:
- Minor corrections: patch version, log
- Factual errors: minor version, notify aggregators
- Major revisions: new major version, archive old, publish erratum

7. Creator Attribution:
- ORCIDs for all creators (persistent IDs)
- Roles clearly defined

Key Insight: Tier 3 metadata strategy = publication-ready documentation. Every field needs provenance, every decision needs policy.

Example 4: Very Complex - Standards Conflict (Workshop Required)

Summary: PROJ-2024-198, 347 pottery fragments, standards conflict → stakeholder workshop

Timeline
- **Nov 25:** Attempt 1 → ❌ Insufficient Plan (return to Phase 2)
- **Dec 10:** Stakeholder workshop
- **Dec 20:** Attempt 2 (revised) → ✅ Approved

Attempt 1 - Conflicting Standards

The Problem:
Museum trying to maintain THREE parallel metadata systems:
1. MARC21 (for library catalog)
2. LIDO (for web)
3. Dublin Core (for repository)

Plus:
- Legacy vocabulary (1970s, unmapped)
- "Multiple versions in different systems" = no single source of truth
- 347 objects = unsustainable to maintain 3× metadata

Critical Decision: Return to Phase 2

Why not just "Needs Rework":
This isn't missing details - it's a fundamentally unsustainable approach. Team will spend years maintaining drifting metadata across 3 incompatible systems.

Reviewer's recommendation: Stakeholder workshop to select ONE primary standard.

The Workshop (Dec 10)

Participants:
- Museum curator
- Library system manager

- Metadata consultant
- IT department
- Collections staff

Key Decisions:
1. **LIDO as single primary standard** (authoritative)
2. **Automated crosswalks** to MARC21/DC/EDM (not maintained manually)
3. **Phased vocabulary migration:** Legacy → Getty AAT over 6 months
4. **Single source of truth:** LIDO master, others derived
5. **Collection-level approach:** 347 objects = collection metadata, detailed for 20 exemplars only

Insight: Standards conflict requires ORGANIZATIONAL decision, not just technical solution.

Attempt 2 - Unified Strategy (Approved!)

Sustainable Approach:

1. Single Primary Standard:
- LIDO (master/authoritative)
- MARC21/DC generated automatically nightly
- No manual maintenance of derived formats

2. Vocabulary Migration:
- Phase 2A: Use legacy with mapping table (380/450 terms mapped)
- Phase 2B: Full Getty AAT migration (6 months)
- During transition: both terms recorded

3. Version Control:
- Git-based for LIDO XML
- Every change tracked (who, when, why)
- Legacy MARC archived (read-only)

4. Collection-Level Strategy:
- 347 fragments = assemblage (research value in patterns)
- Collection metadata for all
- Detailed Tier 3 metadata for 20 exemplars
- Realistic scale

5. Automated Export:
- LIDO updated → all formats regenerated

- Reduces maintenance burden 3x
- Single point of update

Result: Sustainable for decades, not just months.

Comparison Table

Aspect	Example 1	Example 2	Example 3	Example 4
Object	Ceramic Vase	Sculpture	Castle Ruins	347 Pottery Fragments
Tier	1	2	3	3
Standard	Dublin Core	LIDO + vocabularies	LIDO + EDM + multilingual	LIDO (unified)
Attempts	1	1	3	2 + workshop
Days	1	1	8	26
Main Issue	None	None	Missing Tier 3 details	Standards conflict
Resolution	Immediate	Immediate	Iterative detail addition	Organizational workshop
Key Lesson	Keep simple	Controlled vocabs matter	Tier 3 = provenance	One primary standard

---

General Patterns

Success Factors
1. **Match standard to purpose** (DC for discovery, LIDO for research)
2. **Document relationships** (if multiple standards, which is primary?)
3. **Use persistent IDs** (URIs, not just human-readable names)
4. **Plan for updates** (metadata isn't static)

Common Tier 3 Mistakes
1. **Vague multilingual strategy** ("English and Polish" - who translates?)
2. **Undocumented local vocabularies** (no URI, no version)
3. **Missing crosswalks** ("EDM for aggregation" - how? which fields?)
4. **No update policy** (what if errors found?)
5. **Ignoring metadata rights** (metadata itself needs license)

When to Return to Phase 2
- Multiple incompatible primary standards (can't maintain parallel systems)
- Legacy systems with no migration plan
- Organizational disagreement on approach

- Collection-level vs object-level confusion

Time Estimates

- **Tier 1:** 1 day (standard selection straightforward)
- **Tier 2:** 1-2 days (vocabulary selection takes time)
- **Tier 3:** 3-8 days (multilingual, aggregation, rights, updates)
- **Tier 3 with reworks:** +2-3 days per iteration
- **Standards conflict resolution:** 20-30 days (includes workshop)

Recommendations

For Tier 1:
- Dublin Core Simple usually sufficient
- Don't over-engineer

For Tier 2:
- Use established controlled vocabularies (Getty AAT, not invented terms)
- Document vocabulary source URLs
- If bilingual, specify which is authoritative

For Tier 3:
- Every controlled term needs persistent ID (URI)
- Translation workflow with QC specified
- Complete crosswalk tables (not just "we'll use EDM")
- Update policy for post-publication corrections
- Metadata rights (CC0 recommended for interoperability)
- Creator ORCIDs for attribution

For Standards Selection:
- ONE primary standard (authoritative)
- Others derived/exported (not maintained separately)
- If legacy migration: phased approach with mapping table
- Collection-level for large assemblages (not 300+ individual records)

Document Version: 1.0
Date: 2026-01-16

Quick Guide

Phase 2: Metadata Strategy - Quick Guide

🎯 For Evaluators

This package contains:
- **BPMN Diagram:** `02_Phase2_MetadataStrategy_2.bpmn`
- **Example Data:** `phase2_examples_data.xml` (4 projects, 8 submissions)
- **Interpretations:** `phase2_interpretation.md` - Including standards conflict resolution
- **Validation Checklist:** `phase2_validation_checklist.md`
- **This Guide:** Quick reference

Quick Start: Read Example 4 interpretation (standards conflict scenario).

📋 Overview

Purpose: Define HOW you'll describe your digital objects
Duration: 1-8 days depending on tier
Output: Metadata Strategy Paradata Record (XML)

Critical: Metadata decisions affect discoverability for decades. Get this right.

🎯 What This Phase Does

Phase 2 answers:
- What metadata standard? (Dublin Core? LIDO? MARC21?)
- What language(s)?
- What controlled vocabularies? (Getty AAT? Local terms?)
- How will metadata be versioned and updated?
- (Tier 3) How aggregate to Europeana/repositories?

🏊 Three-Tier Structure

Tier 1: Minimum Viable

Tasks:
1. Select Metadata Standard (typically Dublin Core for simple discovery)
2. Define Language of Description

Example: Museum vase, online catalog → Dublin Core (Polish)

Tier 2: Scientific
Additional Tasks:
3. Record Metadata Versioning (track changes over time)
4. Select Thesauri/Vocabularies (Getty AAT, GeoNames, etc.)
5. Record Source of Metadata (card catalog? field notes?)
6. Record Metadata Creator Role (curator vs. metadata specialist)

Example: Research sculpture → LIDO + Getty AAT + bilingual

Tier 3: Memory Twin
Additional Tasks:
7. Define Aggregation Standard (EDM for Europeana, etc.)
   - Must include **complete crosswalk table** (LIDO → EDM field mapping)
8. Document translation workflow (if multilingual)
9. Define metadata update policy
10. Specify metadata rights strategy

Example: Castle ruins → LIDO + EDM + 3 languages + update policy

---

🚀 How to Use

Step 1: Select Primary Standard
Match standard to purpose:
- Discovery/outreach → Dublin Core
- Museum objects/research → LIDO
- Library materials → MARC21
- Archives → EAD

Critical: If multiple standards needed, pick ONE as primary (authoritative). Others derived automatically.

Step 2: Controlled Vocabularies (Tier 2+)

Use established thesauri:
- Materials/techniques: Getty AAT
- Places: GeoNames, TGN
- Periods: PeriodO
- **Don't invent terms** - map to standards

Step 3: Document Everything (Tier 3)
- Translation workflow with QC
- Complete crosswalk tables (not "we'll use EDM")
- Update policy (minor/major/factual corrections)
- Metadata rights (CC0 recommended)

Step 4: Approval
- ✅ **Approved** → Phase 3
- ⚠️ **Needs Rework** → Fix details
- 🔙 **Return to Phase 2** → Standards conflict, need workshop

---

📊 Decision Points

Gateway 1: "Tier ≥ 2?"
If Tier 1: Done (just standard + language)
If Tier 2+: Continue to versioning/vocabularies

Gateway 2: "Tier = 3?"
If Tier 2: Done
If Tier 3: Continue to aggregation strategy

💾 Output Example (Tier 3)

```xml
<MetadataStrategy phase="2">
  <Tier1Data>
    <MetadataStandard>LIDO v1.1 (primary)</MetadataStandard>
    <Language>Polish (pl) primary, English (en) full translation</Language>
  </Tier1Data>

  <Tier2Data>
    <Versioning>v2.0, created 2025-01-23</Versioning>
    <Vocabularies>
      <Material>Getty AAT (with URIs)</Material>
```

```
      <Geographic>GeoNames ID: 3088825</Geographic>
    </Vocabularies>
    <Creators>
      <Curator>Prof. Kamiński (ORCID: 0000-0001-2345-6789)</Curator>
      <MetadataSpecialist>Dr. Lewandowska</MetadataSpecialist>
    </Creators>
  </Tier2Data>

  <Tier3Data>
    <Aggregation>
      <Standard>EDM v5.2.8 for Europeana</Standard>
      <Crosswalk>LIDO descriptive → EDM Proxy</Crosswalk>
      <Documentation>https://iae.edu.pl/lido-edm-crosswalk.pdf</Documentation>
    </Aggregation>
    <TranslationWorkflow>
      Polish (curator) → English (translator) → QC (curator review)
    </TranslationWorkflow>
    <UpdatePolicy>
      Minor: patch version. Major: new version, archive old.
    </UpdatePolicy>
    <MetadataRights>
      <License>CC0 1.0 (metadata), CC BY 4.0 (text)</License>
    </MetadataRights>
  </Tier3Data>
</MetadataStrategy>
```

⚠️ Common Mistakes

Standards Proliferation
❌ **Don't:** Maintain LIDO + MARC + DC all as separate primary systems
✅ **Do:** LIDO primary, MARC/DC exported automatically

Vague Multilingual
❌ **Don't:** "English and Polish"
✅ **Do:** "Polish authoritative, English full translation by certified translator, QC by curator"

Missing Crosswalks
❌ **Don't:** "We'll aggregate to EDM"
✅ **Do:** "LIDO lido:descriptiveMetadata → EDM ore:Proxy. Full crosswalk: [URL]"

Undocumented Local Terms
❌ **Don't:** "Internal vocabulary"
✅ **Do:** "Local vocab v2023.1, URI: https://institution.edu/vocab/"

🔗 Related Phases

- **Previous:** Phase 1 (Planning)
- **Next:** Phase 3 (Data Acquisition)
- **May return from:** Phase 3, 4, 5 if metadata issues discovered

---

📞 Standards Referenced

- Dublin Core Metadata Initiative (DCMI)
- LIDO v1.1 (museum objects)
- EDM (Europeana Data Model)
- Getty Vocabularies (AAT, TGN)
- RightsStatements.org
- ORCID (creator IDs)

Version: 1.0
Last Updated: 2026-01-16

📊 Example Scenarios Comparison

Aspect	Example 1	Example 2	Example 3	Example 4
Object	Ceramic Vase	Sculpture	Castle Ruins	347 Pottery Fragments
Tier	1 - Minimum	2 - Scientific	3 - Memory Twin	3 - Memory Twin
Standard	Dublin Core	LIDO + vocabularies	LIDO + EDM + multilingual	LIDO (unified from 3)
Complexity	⭐ Simple	⭐⭐ Simple	⭐⭐⭐ Complex	⭐⭐⭐⭐ Standards conflict
Attempts	1	1	3	2 + workshop
Days	1	1	8	26
Main Challenge	None	None	Missing Tier 3 details (2x)	3 parallel standards

| **Resolution** | Immediate | Immediate | Iterative detail addition | Stakeholder workshop |
| **Key Lesson** | Keep simple | Use established vocabs | Tier 3 = provenance | One primary standard |

Scenarios Explained:

Example 1 & 2 (Simple):
- Clear purpose, appropriate standard
- Approved immediately
- Templates for straightforward projects

Example 3 (Complex - Multilingual):
- 2 reworks to add Tier 3 detail
- Translation workflow, crosswalks, update policy all needed
- Shows depth of Tier 3 metadata documentation

Example 4 (Very Complex - Standards Conflict):
- Attempting 3 parallel primary standards = unsustainable
- Stakeholder workshop to select ONE primary (LIDO)
- Others automated/derived
- Shows when organizational intervention needed

For detailed analysis, see `phase2_interpretation.md`

Validation

Validation Checklist - Phase 2 Metadata Strategy

Purpose
Validate that metadata strategy is complete, sustainable, and appropriate for tier before Phase 3.

TIER 1: Minimum Viable

Required Fields
- [ ] Metadata standard specified (e.g., "Dublin Core Simple")
- [ ] Language of description specified (e.g., "Polish (pl)")

- [ ] Standard appropriate for purpose (discovery/outreach typically DC)

Red Flags:
- ⚠️ Multiple primary standards for simple project
- ⚠️ Overly complex standard for Tier 1 purpose

Decision: Tier 1 complete? **YES / NO / NEEDS REWORK**

TIER 2: Scientific

Metadata Versioning
- [ ] Version history documented (initial creation, updates)
- [ ] Migration from legacy documented (if applicable)
- [ ] Version numbers assigned

Thesauri/Vocabularies
- [ ] Controlled vocabularies specified (Getty AAT, GeoNames, etc.)
- [ ] Source URLs or identifiers provided
- [ ] Appropriate for domain (art history ≠ archaeology terms)

Metadata Source
- [ ] Original sources documented (card catalog, field notes, etc.)
- [ ] Integration of multiple sources explained

Creator Roles
- [ ] Specific individuals named (not just "team")
- [ ] Roles defined (curator vs. metadata specialist vs. translator)

Warning Signs:
- ⚠️ "Internal vocabulary" without documentation
- ⚠️ Vague source ("various catalogs")
- ⚠️ Generic roles ("metadata team")

Decision: Tier 2 complete? **YES / NO / NEEDS REWORK**

TIER 3: Memory Twin

Aggregation Standard
- [ ] Target aggregator specified (Europeana, DPLA, etc.)
- [ ] Aggregation standard specified (EDM, MARC21, etc.)

- [ ] **Crosswalk documented:** Primary standard → Aggregation format
- [ ] Crosswalk detailed (field-level mapping, not just "we'll use EDM")

Translation Workflow (if multilingual)
- [ ] Translation process documented (who, when, how)
- [ ] Quality control specified
- [ ] Timeline provided
- [ ] Primary/authoritative language identified

Metadata Update Policy
- [ ] Policy for minor corrections (typos)
- [ ] Policy for factual errors
- [ ] Policy for major revisions
- [ ] Change log location specified

Metadata Rights Strategy
- [ ] License for metadata specified (CC0 recommended)
- [ ] License for descriptive text
- [ ] License for digital objects
- [ ] Compliance with RightsStatements.org noted

Creator Attribution
- [ ] Persistent IDs for creators (ORCIDs recommended)
- [ ] Clear roles and responsibilities

Red Flags for Tier 3:
- 🚩 Multiple primary standards (LIDO + MARC + DC all maintained separately)
- 🚩 "We'll translate" without workflow
- 🚩 Aggregation mentioned without crosswalk
- 🚩 No update policy
- 🚩 No metadata rights statement
- 🚩 Local vocabulary without URI/version

Decision: Tier 3 complete? **YES / NO / NEEDS REWORK**

CROSS-TIER CHECKS

Standards Sustainability
- [ ] If multiple standards: ONE is primary, others derived/exported
- [ ] No plan to maintain 2+ standards manually in parallel
- [ ] Legacy vocabulary mapped to modern standards (or migration plan)

Collection vs. Object Level
- [ ] For 50+ objects: collection-level strategy considered?
- [ ] If collection: exemplars identified for detailed metadata?

Controlled Vocabulary Quality
- [ ] Terms from established thesauri (Getty AAT, GeoNames, not invented)
- [ ] Persistent IDs (URIs) for Tier 3
- [ ] Local vocabularies documented with version and URI

APPROVAL DECISION MATRIX

Situation	Decision	Action
All requirements met	✅ **Approved**	Proceed to Phase 3
Missing documentation, fixable	⚠️ **Needs Rework**	Specific feedback, Phase 2
Multiple incompatible primary standards	🔙 **Return to Phase 2**	Stakeholder workshop
Legacy system, no migration plan	🔙 **Return to Phase 2**	Develop migration strategy
Collection scale issue (100s of objects)	🔙 **Return to Phase 2**	Collection-level approach

---

WHEN TO RETURN TO PHASE 2 START

Use "Return to Phase 2" for:
1. **Standards conflict:** Attempting to maintain 2+ primary standards in parallel
2. **Legacy migration absent:** Old system, no plan to integrate/replace
3. **Organizational disagreement:** Different departments want different standards
4. **Scale mismatch:** Individual object metadata for 300+ objects (need collection approach)

Requires stakeholder workshop to resolve.

FEEDBACK EXAMPLES

Needs Rework (Tier 3)

> "NEEDS REWORK: Multilingual strategy incomplete. You state 'Polish and English' but don't specify: (1) which is authoritative, (2) translation workflow (who translates?), (3) quality control process, (4) timeline. EDM aggregation mentioned but crosswalk table missing - need field-level mapping (LIDO element → EDM property). Missing metadata update policy - what happens if factual error discovered post-publication? Estimate 3-4 days to complete."

Return to Phase 2
> "RETURN TO PHASE 2: Attempting to maintain three parallel primary standards (MARC21 + LIDO + Dublin Core) is unsustainable - metadata will drift across systems. Recommend stakeholder workshop to: (1) Select ONE primary standard (suggest LIDO for museum objects), (2) Define automated crosswalks to other formats, (3) Establish single source of truth. Current approach will require perpetual manual synchronization = technical debt. Also: 347 objects need collection-level strategy, not 347 individual Tier 3 records."

Approved (Tier 3)
> "APPROVED: Exemplary Tier 3 metadata strategy. Single primary standard (LIDO) with automated exports. Complete LIDO→EDM crosswalk documented with URL. Multilingual workflow specified with QC and timeline. All vocabularies use persistent URIs (Getty AAT IDs, GeoNames IDs). Comprehensive update policy (minor/major/factual). Metadata rights: CC0 (excellent for interoperability). ORCIDs for all creators. Ready for Phase 3."

TIME ESTIMATES

- **Tier 1:** 5-10 min review
- **Tier 2:** 15-20 min review
- **Tier 3:** 30-45 min (detailed crosswalk/workflow review)
- **Re-review:** 10-15 min

COMMON PITFALLS

1. **Standards proliferation:** Trying to maintain multiple primary standards
2. **Vague multilingual:** "We'll translate" (no workflow)
3. **Missing crosswalks:** "We'll use EDM" (how? which fields?)
4. **Undocumented local terms:** No URI, no version control
5. **No update policy:** Metadata treated as immutable
6. **Metadata rights ignored:** Only digital object rights documented

---

FINAL CHECKLIST

- [ ] Tier requirements reviewed
- [ ] Standards sustainability checked (not maintaining parallel systems)
- [ ] For Tier 3: crosswalks, workflow, updates, rights all documented
- [ ] Controlled vocabularies appropriate and persistent (URIs for Tier 3)
- [ ] Collection vs. object level appropriate
- [ ] Feedback specific and actionable

Reviewer: _______________
Date: _______________
Decision: Approved / Needs Rework / Return to Phase 2

Version: 1.0
Last Updated: 2026-01-16

13.3.3 Phase 3: Data Acquisition

Interpretation

Phase 3 Interpretation
Example 1-2: Simple captures, straightforward
Example 3: Castle - 2 re-captures for lighting/coverage issues. Shows Tier 3 field complexity.
Example 4: EQUIPMENT CRISIS - Scanner failed mid-project (180/347 complete). Workshop → hybrid methodology (structured light + photogrammetry). Cross-validation study (20 objects both methods) validates approach. Demonstrates adaptive response while maintaining scientific rigor.
Key lesson: Tier 3 needs equipment redundancy planning. Single point of failure = project risk.
Version: 1.0 | 2026-01-17

Quick Guide

Phase 3: Data Acquisition - Quick Guide

🎯 For Evaluators

Package contains:
- **BPMN:** `03_Phase3_DataAcquisition_1.bpmn`
- **Examples:** `phase3_examples_data.xml` (4 projects)
- **Interpretations:** `phase3_interpretation.md` (equipment failure scenario)
- **Checklist:** `phase3_validation_checklist.md`
- **This Guide**

Key: Example 4 = equipment crisis response

📋 Overview

Purpose: Physically capture digital data
Duration: 1 day - 5 weeks
Output: Raw dataset + Acquisition Paradata

Critical: Quality issues here can't be fixed in Phase 4. Get capture right.

🏊 Three-Tier Structure

Tier 1: Minimum Viable
Tasks:
1. Record Acquisition Method (photogrammetry/scan/2D)
2. Record Equipment Hardware (camera/scanner model)
3. Record Operator Name
4. Record Capture Quality Metric (resolution/GSD/point density)

Example: Museum vase, 85 images, 20MP camera

Tier 2: Scientific
Additional:
5. Record Quality Assessment Standard (FADGI/Metamorfoze)
6. Record Dataset Raw Size (GB, image count)

Example: Sculpture, FADGI 3-star, 252 images, 12.6GB

Tier 3: Memory Twin
Additional (Manual tasks - field work):
7. Record Metric Calibration (scale bars/coded targets/GCPs)
8. Record Lighting Setup (natural/artificial/cross-polarized)
9. Record Color Calibration (ColorChecker placement/protocol)
10. Record Environmental Conditions (weather/settings/logs)

Example: Castle, RTK-GPS survey, 5-day weather log, HDR bracketing

---

📊 Example Scenarios

Aspect	Ex 1	Ex 2	Ex 3	Ex 4
Object	Vase	Sculpture	Castle	347 Pottery
Tier	1	2	3	3
Method	Photos	Turntable	Drone+Ground	Scanner→Photos
Attempts	1	1	3	2
Issue	None	None	Shadows, coverage gaps	Equipment failure
Resolution	-	-	Re-capture 2×	Hybrid methodology
Lesson	Simple works	Calibration matters	Conditions critical	Redundancy needed

For full analysis: `phase3_interpretation.md`

Version: 1.0 | 2026-01-17

Validation

Phase 3 Validation Checklist
Tier 1: Method, equipment, operator, quality metric
Tier 2: + Standard (FADGI/Metamorfoze), dataset size
Tier 3: + Calibration (scale bars/GCPs), lighting (protocol), color (ColorChecker), environment (weather logs)
Red flags: No calibration, vague lighting, missing environmental docs
Equipment failure: Approve if documented response + methodology validation
Version: 1.0 | 2026-01-17

13.3.4 Phase 4: Processing and Enrichment

Interpretation

Examples Interpretation - Phase 4 Processing & Enrichment

Overview

This document explains 4 real-world scenarios of Phase 4 execution, from simple to very complex:
- **Example 1 & 2:** Simple, straightforward (approved immediately)
- **Example 3 & 4:** Complex, with multiple reworks and learning opportunities

---

Example 1: Simple Tier 1 - Greek Ceramic Vase

Project Summary
- **ID:** PROJ-2025-001
- **Object:** Museum Artifact - Greek Ceramic Vase
- **Institution:** City Museum of Archaeology
- **Tier:** 1 (Minimum Viable)
- **Complexity:** Simple

Timeline
- **Jan 10, 2025:** Submitted
- **Jan 11, 2025:** Approved ✅

What Happened
This is a **textbook example** of a simple Tier 1 project done right the first time.

Key Decisions:
- Used **Meshroom** (free, open-source) - appropriate for museum's budget
- Target: Web visualization only (no research requirement)
- Optimization: 96% polygon reduction (2.5M → 100K)
- Processing straightforward, no enrichment needed

Why It Worked:
1. ✅ Clear, limited scope (web display only)
2. ✅ Appropriate software choice
3. ✅ Complete Tier 1 documentation

4. ✅ No unnecessary complexity

Result: Approved in 1 day. Total project time: 2 days.

Lesson: When scope is clear and limited, Tier 1 is fast and efficient.

Example 2: Simple Tier 2 - Monument Statue

Project Summary
- **ID:** PROJ-2024-156
- **Object:** General Władysław Sikorski Statue (Public Monument)
- **Institution:** National Heritage Institute
- **Tier:** 2 (Scientific)
- **Complexity:** Simple

Timeline
- **Dec 18, 2024:** Submitted
- **Dec 20, 2024:** Approved ✅

What Happened
A **well-executed Tier 2 project** with proper planning.

Key Decisions:
- Used **Metashape Professional** - industry standard for research
- Target: Historical documentation + research
- LOD system for different use cases (5M/2M/500K/100K)
- High-quality capture: 250 photos, DSLR camera
- **Tier 2 additions:** Processing parameters, uncertainty assessment

Documentation Quality:
- ✅ All processing parameters recorded (alignment, depth filtering, etc.)
- ✅ Uncertainty level justified: "4" (high confidence, minor uncertainty in shadows)
- ✅ 45 million point cloud size documented

Why It Worked:
1. ✅ Team experienced with Tier 2 requirements
2. ✅ Good capture quality = less processing issues
3. ✅ Clear documentation from the start

Result: Approved in 2 days. Total project time: 5 days.

Lesson: Tier 2 requires more documentation, but with good planning and capture quality, it's still straightforward.

Example 3: Complex Tier 3 - Medieval Manuscript (3 Attempts)

Project Summary
- **ID:** PROJ-2025-003
- **Object:** Medieval Manuscript - Codex Illuminatus (15th century)
- **Institution:** National Library
- **Tier:** 3 (Memory Twin)
- **Complexity:** Complex

Timeline
- **Jan 10:** Attempt 1 submitted → ❌ Needs Rework
- **Jan 14:** Attempt 2 submitted → ❌ Needs Rework
- **Jan 16:** Attempt 3 submitted → ✅ Approved

Total time: 6 days, 3 submissions

Attempt 1 - Needs Rework (Jan 10)

What Was Submitted:
- Processing software mentioned but incomplete versions
- Brief mention of OCR corrections - no details
- AI model: "BERT mentioned but no version"
- Manual intervention log: vague

Rejection Reason:
> "Processing parameters incomplete (missing surface reconstruction settings, texture resolution). Manual intervention log lacks dates, operator names, and specific corrections. AI model documentation insufficient - no version, no source, no accuracy metrics."

Problems:
1. ❌ Incomplete Tier 2 documentation (processing params)
2. ❌ Tier 3 manual intervention log too vague
3. ❌ AI model barely documented

Team's Mistake: Assumed "mentioning" AI model was enough. Didn't realize Tier 3 requires full provenance.

Attempt 2 - Still Needs Rework (Jan 14)

What Changed:
- ✅ Added exact software versions
- ✅ Completed all Tier 2 processing parameters
- ✅ Improved manual intervention log (dates, operator, specific corrections)
- ⚠️ AI model: added version and source, but still incomplete

Rejection Reason:
> "Better, but AI model still incomplete. Missing: training dataset description, accuracy/validation metrics (F1-score, precision, recall), license information. Manual intervention log improved but needs second intervention (metadata cleanup mentioned in Phase 3 notes)."

Problems:
1. ❌ AI model documentation still insufficient
2. ❌ Missing second manual intervention

Team's Realization: Tier 3 AI documentation must be **publication-ready** - training data, metrics, reproducibility.

Attempt 3 - Approved! (Jan 16)

What Changed:
- ✅ Added ALL AI model details:
  - Training data: "Medieval Latin corpus (15th-16th century, 2.3M tokens)"
  - Metrics: "F1-score: 89.3%, Precision: 91.2%, Recall: 87.5%"
  - License: Apache 2.0
  - Source URL: Hugging Face link
- ✅ Added second manual intervention (metadata cleanup)
- ✅ Time spent documented for each intervention

Approval Comment:
> "All Tier 3 documentation now complete. Processing parameters comprehensive. Manual intervention log detailed with dates, operators, and time spent. AI model fully documented with training data, validation metrics, and license. Ready for Phase 5."

Result: ✅ Approved

Key Lessons from Example 3

What Went Wrong:
- Team underestimated Tier 3 documentation requirements
- Initially treated AI model as "just a tool" rather than part of the research method
- Manual interventions not tracked in real-time

What Went Right:
- Team responded quickly to feedback (2-day turnaround per iteration)
- Quality Control reviewer gave **specific, actionable** feedback
- Each iteration improved significantly

Cost of Rework:
- Time: +4 days vs. if done right first time
- Effort: 3 submissions instead of 1
- BUT: Final quality excellent, documentation publication-ready

Takeaway:
> **Tier 3 is research documentation.** Every AI model, every manual edit, every parameter must be **reproducible** and **auditable**. If you can't publish it in a methods section, it's not documented enough.

Example 4: Very Complex Tier 3 - Archaeological Ceramic (Phase 3 Return)

Project Summary
- **ID:** PROJ-2024-125
- **Object:** Ceramic Vessel Fragment (Bronze Age)
- **Institution:** Institute of Archaeology
- **Tier:** 3 (Memory Twin)
- **Complexity:** Very Complex

Timeline
- **Nov 20:** Attempt 1 → ❌ **Insufficient Plan** (return to Phase 3)
- **Nov 22:** Decision to return to Phase 3
- **Dec 5:** Re-acquisition in Phase 3 (improved setup)
- **Dec 10:** Attempt 2 → ❌ Needs Rework
- **Dec 15:** Attempt 3 → ✅ Approved
- **Dec 16:** Final approval

Total time: 26 days (including re-acquisition)

Attempt 1 - Insufficient Plan (Nov 20)

What Was Submitted:
- Tier 1: Standard processing
- Tier 2: "Standard Metashape High quality settings" (vague)
- Uncertainty: **Geometry: 2, Texture: 2** (very low!)
- AI model: "Custom CNN v2.1" - no other details
- Manual interventions: "Multiple corrections made - details not recorded"

Critical Decision: Return to Phase 3

Reviewer's Analysis:
> "CRITICAL ISSUES: 1) Uncertainty level too low (Geometry: 2) suggests fundamental acquisition problems - likely poor lighting. 2) AI model completely undocumented. 3) Manual intervention log shows 'multiple corrections' but no specifics - red flag for data integrity. 4) RTI enrichment mentioned but no RTI-specific parameters. RECOMMENDATION: Return to Phase 3 with improved lighting setup."

Why This Was The Right Decision:
- Uncertainty level of "2" = **NOT FIT FOR PURPOSE** (Tier 3 requires high confidence)
- Trying to fix in Phase 4 = polishing bad data
- Better to re-acquire properly than document failure

Required Actions Before Returning:
1. Re-capture with cross-polarized + diffuse lighting
2. Add color calibration (X-Rite ColorChecker)
3. Document RTI setup
4. Prepare complete AI model documentation

Re-Acquisition in Phase 3 (Dec 5)

Improvements Made:
- **Lighting:** Cross-polarized filters + LED dome (64 positions) - eliminated specular highlights
- **Color:** X-Rite ColorChecker Passport for calibration
- **RTI:** Full hemispherical capture (64 light positions) - reveals micro-topology
- **Photos:** 380 images (up from 180) - better coverage

Cost of Phase 3 Return:

- Time: +15 days
- Equipment: Additional lighting gear
- Expertise: RTI specialist brought in
- BUT: Data quality dramatically improved

Attempt 2 - After Phase 3 Return (Dec 10)

What Changed:
- ✅ Much better data quality
- ✅ Uncertainty now "4-5" (acceptable for Tier 3)
- ✅ RTI parameters documented
- ✅ AI model better described (architecture, purpose)
- ❌ Still missing: training dataset details, validation metrics, manual intervention log

Rejection Reason:
> "Much improved after Phase 3 return. Quality now acceptable. However, AI model documentation still incomplete: Missing training dataset description (how many samples?), missing validation metrics (accuracy, F1-score), missing information about model weights availability. Manual intervention log still not provided."

Team's Progress: Data quality problem solved, but documentation still catching up.

Attempt 3 - Finally Approved (Dec 15)

What Changed:
- ✅ Complete AI model documentation:
  - Training: 2,847 annotated Bronze Age pottery images from 12 sites
  - Metrics: 87.4% accuracy, F1: 85.9%
  - Architecture: ResNet50 transfer learning
  - Availability: Model weights on request
- ✅ Manual intervention log:
  - Pattern annotation: 4 hours by ceramics specialist
  - RTI alignment corrections: 1.5 hours
- ✅ Enrichment methodology: "Hybrid - automated detection with expert validation"

Approval Comment:
> "Excellent turnaround after Phase 3 return. Re-acquisition significantly improved data quality. All Tier 3 documentation now comprehensive. This project demonstrates the value of returning to earlier phases when quality issues are identified. Ready for Phase 5."

Result: ✅ Approved

Key Lessons from Example 4

When to Return to Earlier Phase:
- Uncertainty level < 3 in Tier 3 = red flag
- "Multiple corrections" without log = data integrity concern
- Fix root cause (acquisition) rather than symptoms (processing)

Cost-Benefit of Phase 3 Return:

Aspect	If Continued	With Return
Time	~10 days (estimated)	26 days
Quality	Low (uncertainty: 2)	High (uncertainty: 4-5)
Research Value	Questionable	Excellent
Publication Ready	No	Yes

Team's Growth:
- Learned that **quality can't be fixed in post-processing**
- Invested in proper RTI equipment
- Developed comprehensive AI documentation template
- Now have reusable lighting setup for future projects

Takeaway:
> **Don't be afraid to go back.** Returning to Phase 3 added 15 days but resulted in research-grade data. Continuing with poor acquisition would have produced a "Tier 3" dataset that wasn't fit for purpose.

Comparison Table

Aspect	Example 1	Example 2	Example 3	Example 4
Object	Ceramic Vase	Monument Statue	Medieval Manuscript	Bronze Age Pottery
Tier	1	2	3	3
Complexity	Simple	Simple	Complex	Very Complex
Submissions	1	1	3	3 (+ Phase 3 return)
Total Days	2	5	6	26

Software	Meshroom (OSS)	Metashape Pro	RealityCapture + ABBYY	Metashape + RTIBuilder
AI Used	No	No	Yes (BERT)	Yes (Custom CNN)
Enrichment	None	None	OCR + False Color	RTI + Pattern Detection
Main Challenge	None	None	AI documentation	Poor initial capture
Key Learning	Keep it simple	Plan well	Document AI thoroughly	Fix acquisition first
Final Status	✅ Approved	✅ Approved	✅ Approved (after 2 reworks)	✅ Approved (after Phase 3)

General Patterns Observed

Success Factors
1. **Clear scope definition** (Examples 1-2)
2. **Good capture quality** → Less processing issues (Example 2)
3. **Responsive to feedback** (Example 3 - quick iterations)
4. **Willingness to go back** (Example 4 - quality over speed)

Common Mistakes
1. **Underestimating Tier 3 documentation** (Examples 3-4)
2. **Treating AI as "black box"** (Examples 3-4)
3. **Not logging interventions in real-time** (Examples 3-4)
4. **Trying to fix acquisition problems in processing** (Example 4 - initial attempt)

Time Estimates by Tier
- **Tier 1:** 2-3 days (if straightforward)
- **Tier 2:** 5-7 days (with proper documentation)
- **Tier 3:** 6-10 days (first time), 3-5 days (with experience)
- **Tier 3 + reworks:** Add 2-4 days per iteration
- **Phase 3 return:** Add 15-20 days for re-acquisition

Recommendations for Future Projects

For Tier 1:
- Use checklist to ensure all fields filled
- Don't over-document - Tier 1 is intentionally minimal

For Tier 2:
- Export processing parameters **during** processing (don't rely on memory)
- Document uncertainty **honestly** - "4" is better than false "5"

For Tier 3:
- **Before starting:** Prepare AI documentation template
- **During processing:** Log every manual intervention immediately
- **Before submission:** Review with domain expert (not just technical team)
- **AI models:** Treat like publishing a paper - training data, metrics, availability

For Projects with AI:
Create a standard template including:
- Model name/version
- Architecture description
- Training dataset (size, source, period)
- Validation metrics (accuracy, F1, precision, recall)
- Source/weights availability
- License
- Purpose and limitations

For Quality Control:
- Don't approve "Tier 3" with uncertainty < 3
- If you see "multiple corrections" without log → reject
- If AI mentioned without metrics → needs rework
- Phase 3 return is OK - use it when needed

Document Version: 1.0
Date: 2026-01-16
Purpose: Evaluation package for Phase 4 BPMN process

Quick Guide

Phase 4: Processing & Enrichment - Quick Guide

🎯 For Evaluators

This package contains:
- **BPMN Diagram:** `04_Phase4_ProcessingEnrichment.bpmn` - Process flow with 3-tier structure
- **Example Data:** `phase4_examples_data.xml` - 4 real-world scenarios (2 simple, 2 complex with reworks)

- **Interpretations:** `examples_interpretation.md` - Detailed analysis of each example
- **Validation Checklist:** `validation_checklist.md` - How to review Phase 4 submissions
- **This Guide:** Quick reference for process usage

Quick Start: Open BPMN diagram in Camunda Modeler, then read Example 3 interpretation to understand rework scenarios.

📋 Overview

Purpose: Document all processing and enrichment steps applied to digital assets
Duration: ~2-4 hours depending on tier
Output: Processing Paradata Record (XML)

🎯 What This Phase Does

This phase captures **how** you transformed raw capture data into the final digital asset:
- Software used and settings
- Optimization decisions
- Enrichment methods (OCR, annotations, etc.)
- AI/ML models applied
- Manual interventions

🏊 Three-Tier Structure

Tier 1: Minimum Viable
Who needs this: All projects
Tasks:
1. **Record Processing Software** - What tool did you use? (e.g., RealityCapture, Metashape)
2. **Record Optimization** - How did you reduce file size for web/XR? (e.g., "Decimated to 500K polygons")

Example Data:
```xml
<ProcessingSoftware>RealityCapture 1.3.2</ProcessingSoftware>
```

<Optimization>Decimated to 500K polygons for WebXR</Optimization>
```

---

### Tier 2: Scientific
**Who needs this:** Research projects, scholarly repositories
**Additional Tasks:**
3. **Record Processing Parameters** - Exact settings (alignment quality, filtering, etc.)
4. **Record Enrichment Type** - What enrichment did you add? (OCR, false color, annotations)
5. **Record Uncertainty Level** - How confident are you? (1=Hypothetical, 5=Measured)

**Example Data:**
```xml
<ProcessingParameters>
  <AlignmentQuality>High</AlignmentQuality>
  <DepthFiltering>Aggressive</DepthFiltering>
  <SurfaceReconstruction>Poisson (depth: 12)</SurfaceReconstruction>
</ProcessingParameters>
<UncertaintyLevel>
  <GeometryConfidence>4</GeometryConfidence>
  <ColorConfidence>3</ColorConfidence>
</UncertaintyLevel>
```

Tier 3: Memory Twin
Who needs this: High-fidelity digital preservation, complex projects
Additional Tasks:
6. **Record Manual Intervention Log** - Every manual edit/correction
7. **Record Enrichment Methodology** - Manual vs Automated vs Hybrid?
8. **Record AI Model Used** - Which AI/ML models? (name, version, source)

Example Data:
```xml
<ManualInterventionLog>
  <Intervention>
    <Date>2025-01-10</Date>
    <Type>OCR correction</Type>
    <Description>Corrected 43 misread Gothic characters</Description>
    <Operator>Dr. Anna Kowalska</Operator>
```
```

```xml
  </Intervention>
</ManualInterventionLog>
<AIModelUsed>
  <ModelName>BERT-base-multilingual-cased</ModelName>
  <Purpose>Named Entity Recognition in Latin text</Purpose>
  <Accuracy>89.3% F1-score</Accuracy>
</AIModelUsed>
```

🚀 How to Use

Step 1: Select Your Tier
Based on project requirements (decided in Phase 1)

Step 2: Complete Tasks Sequentially
- Start in **Tier 1** lane (always required)
- If Tier ≥ 2: Move to **Tier 2** lane
- If Tier = 3: Move to **Tier 3** lane

Step 3: Approval
Final check before moving to Phase 5 (Publication)

Approval Options:
- ✅ **Yes** → Proceed to Phase 5
- ❌ **No** → Restart phase (refine data)
- 🔙 **Insufficient** → Go back to Phase 3 (re-acquire data)

📊 Decision Points

Gateway 1: "Tier ≥ 2?"
- **If Tier 1 only** → Skip to Approval
- **If Tier 2 or 3** → Continue to Tier 2 tasks

Gateway 2: "Tier = 3?"
- **If Tier 2 only** → Skip to Approval
- **If Tier 3** → Continue to Tier 3 tasks

💾 Output Format

File: `processing_paradata_[ProjectID].xml`
Schema: Paradata Processing Schema v1.0
Encoding: UTF-8

Example filename: `processing_paradata_PROJ-2025-001.xml`

📝 Example: Complete Tier 3 Record

```xml
<ProcessingEnrichmentData phase="4">
 <Project>
  <ProjectID>PROJ-2025-001</ProjectID>
  <ObjectName>Medieval Manuscript - Codex Illuminatus</ObjectName>
  <SelectedTier>3</SelectedTier>
 </Project>

 <Tier1_MinimumViable>
  <ProcessingSoftware>RealityCapture 1.3.2</ProcessingSoftware>
  <Optimization>Decimated to 500K polygons for WebXR</Optimization>
 </Tier1_MinimumViable>

 <Tier2_Scientific>
  <ProcessingParameters>
   <AlignmentQuality>High</AlignmentQuality>
   <DepthFiltering>Aggressive</DepthFiltering>
  </ProcessingParameters>
  <EnrichmentType>OCR + False Color Imaging</EnrichmentType>
  <UncertaintyLevel>
   <GeometryConfidence>4</GeometryConfidence>
  </UncertaintyLevel>
 </Tier2_Scientific>

 <Tier3_MemoryTwin>
  <ManualInterventionLog>
   <Intervention>
    <Type>OCR correction</Type>
    <Description>Corrected 43 misread characters</Description>
   </Intervention>
  </ManualInterventionLog>
  <EnrichmentMethodology>Hybrid (Human-in-Loop)</EnrichmentMethodology>
  <AIModelUsed>
```

<ModelName>BERT-base-multilingual-cased</ModelName>
<Accuracy>89.3% F1-score</Accuracy>
</AIModelUsed>
</Tier3_MemoryTwin>

<ApprovalRecord>
<Status>Approved</Status>
<Date>2025-01-15</Date>
</ApprovalRecord>
</ProcessingEnrichmentData>
` ` `

⚠️ Common Mistakes

❌ **Don't:** Leave software version blank
✅ **Do:** Record exact version (e.g., "1.3.2.15837")

❌ **Don't:** Say "We optimized the model"
✅ **Do:** Specify method (e.g., "Decimated from 15M to 500K polygons")

❌ **Don't:** Forget manual interventions in Tier 3
✅ **Do:** Log every manual edit with operator name

🔗 Related Phases

- **Previous:** Phase 3 (Data Acquisition)
- **Next:** Phase 5 (Ingest & Publication)
- **References:** Phase 1 (selected tier determines which tasks to complete)

📞 Questions?

Standards Referenced:
- Reproducibility: Documented processing ensures reproducibility
- IMPULSE XR Requirements: Optimization for web/XR delivery
- Seville Principles: Uncertainty documentation
- InterPARES Trust AI: AI model documentation
- Hugging Face Model Cards: AI model metadata

Version: 1.0
Last Updated: 2025-01-16

📊 Example Scenarios Comparison

The evaluation package includes 4 example projects demonstrating different scenarios:

Aspect	Example 1	Example 2	Example 3	Example 4
Object	Greek Vase	Monument Statue	Medieval Manuscript	Bronze Age Pottery
Tier	1 - Minimum	2 - Scientific	3 - Memory Twin	3 - Memory Twin
Complexity	⭐ Simple	⭐⭐ Simple	⭐⭐⭐ Complex	⭐⭐⭐⭐ Very Complex
Attempts	1	1	3	3 + Phase 3 return
Total Time	2 days	5 days	6 days	26 days
Software	Meshroom (OSS)	Metashape Pro	RealityCapture + ABBYY	Metashape + RTI
AI Used	No	No	Yes (BERT NER)	Yes (Custom CNN)
Enrichment	None	None	OCR + False Color	RTI + Pattern Detection
Main Challenge	None	None	AI documentation	Poor initial capture quality
Rework Reason	-	-	Incomplete AI docs (2x)	Phase 3 return + docs
Final Status	✅ Approved	✅ Approved	✅ Approved	✅ Approved
Key Lesson	Keep it simple	Plan documentation early	Document AI like publishing paper	Fix acquisition problems at source

Scenarios Explained:

Example 1 & 2 (Simple - Happy Path):
- Straightforward execution
- Complete documentation from start
- Approved immediately
- **Use these as templates** for standard projects

Example 3 (Complex - Iterative Improvement):
- 2 reworks due to incomplete AI model documentation
- Shows common mistakes: assuming "mentioning" AI is enough
- Demonstrates responsive team fixing issues quickly
- **Use this to understand** rework loops and documentation depth needed

Example 4 (Very Complex - Phase 3 Return):
- Critical decision: return to earlier phase when quality insufficient
- Shows courage to "go back" rather than polish bad data
- Re-acquisition added 15 days but resulted in research-grade output
- **Use this to understand** when "Insufficient Plan" decision is appropriate

For detailed analysis of each scenario, see `examples_interpretation.md`

Validation

Validation Checklist - Phase 4 Processing & Enrichment

Purpose
This checklist helps reviewers validate that Phase 4 documentation is complete and meets tier requirements.

TIER 1: Minimum Viable (Required for ALL projects)

Processing Software
- [ ] Software name recorded
- [ ] Software version number included (e.g., "1.3.2" not just "RealityCapture")
- [ ] Vendor/source identified (commercial, open-source, in-house)

Optimization for User Engagement
- [ ] Optimization method described (decimation, LOD, compression, etc.)
- [ ] Quantitative metrics provided (e.g., "15M → 500K polygons")
- [ ] Target platform specified (web, XR, mobile, etc.)
- [ ] Rationale clear (why this optimization level?)

General
- [ ] All fields have values (no "TBD" or blank)
- [ ] Purpose aligns with project scope from Phase 1

Decision: Tier 1 complete? **YES / NO / NEEDS REWORK**

TIER 2: Scientific (If project tier ≥ 2)

Processing Parameters
- [ ] Key settings documented (alignment quality, filtering, reconstruction method)
- [ ] Settings are **specific** (not "high quality" but "High + Aggressive filtering")
- [ ] Technical details sufficient for reproduction
- [ ] Any custom/unusual settings explained

Enrichment Type
- [ ] Type clearly specified (OCR, RTI, false color, annotations, etc.)
- [ ] Method described (software, technique)
- [ ] Purpose stated (why this enrichment?)
- [ ] "None" explicitly stated if no enrichment applied

Uncertainty Level
- [ ] Documented on 1-5 scale (1=Hypothetical, 5=Measured)
- [ ] Separate ratings if applicable (geometry, color, text, etc.)
- [ ] **Rationale provided** - why this rating?
- [ ] Rating ≥ 3 for Tier 3 projects (if < 3, consider Phase 3 return)

Warning Signs:
- ⚠️ Vague parameters ("standard settings")
- ⚠️ Uncertainty < 3 with no explanation
- ⚠️ Missing rationale

Decision: Tier 2 complete? **YES / NO / NEEDS REWORK**

TIER 3: Memory Twin (If project tier = 3)

Manual Intervention Log
- [ ] ALL manual interventions documented (not just "corrections made")
- [ ] Each intervention includes:
  - [ ] Date (YYYY-MM-DD)
  - [ ] Type (OCR correction, mesh cleanup, alignment, etc.)
  - [ ] Description (specific actions taken)
  - [ ] Operator name and role
  - [ ] (Optional but recommended) Time spent
- [ ] If "no manual interventions" - explicitly state this

Red Flags:
- 🚩 "Multiple corrections made" without details

- 🚩 "Some manual work" without log
- 🚩 Missing operator identification

Enrichment Methodology
- [ ] Approach clearly stated (Manual / Automated / Hybrid)
- [ ] If Hybrid: describes which steps are automated vs. manual
- [ ] If Human-in-the-Loop: workflow explained
- [ ] Time investment documented (human hours, compute hours)

AI Model Used (if applicable)
If AI/ML used, ALL of the following required:

Basic Information
- [ ] Model name
- [ ] Version number
- [ ] Architecture (CNN, Transformer, traditional ML, etc.)
- [ ] Purpose (what task does it perform?)

Provenance
- [ ] Source (Hugging Face, GitHub, commercial, in-house)
- [ ] If pre-trained: original training data mentioned
- [ ] If custom: developed by whom?
- [ ] Code/weights availability status

Training & Validation (Critical!)
- [ ] Training dataset described:
  - [ ] Size (number of samples)
  - [ ] Source/period/domain
  - [ ] Relevance to current object
- [ ] Validation metrics provided:
  - [ ] At minimum: Accuracy OR F1-score
  - [ ] Ideally: Precision, Recall, F1
  - [ ] Validation method (cross-validation, held-out test set)
- [ ] Known limitations discussed

Licensing & Access
- [ ] License specified (Apache, MIT, proprietary, research-only, etc.)
- [ ] Access conditions clear (open, on request, restricted)
- [ ] Citation/attribution requirements noted

Red Flags for AI Documentation:
- 🚩 "AI model used" without details → REJECT
- 🚩 Model name without version → NEEDS REWORK
- 🚩 No training data description → NEEDS REWORK

- 🚩 No validation metrics → NEEDS REWORK
- 🚩 "Custom model" without source code or availability statement → NEEDS REWORK

Special Note:
If AI model is undocumented, the entire enrichment is not reproducible. **This is a blocking issue for Tier 3.**

Decision: Tier 3 complete? **YES / NO / NEEDS REWORK**

CROSS-TIER CHECKS

Consistency
- [ ] Tier matches selected tier from Phase 1
- [ ] Complexity assessment aligns with data (simple objects shouldn't need Tier 3)
- [ ] Processing choices align with acquisition method from Phase 3

Quality Signals
- [ ] No obvious copy-paste errors (e.g., wrong object name)
- [ ] Dates are logical (submission after execution)
- [ ] Technical details are plausible (not impossible values)

Approval Decision Factors
- [ ] If Tier 3 and uncertainty < 3 → Consider **Insufficient Plan** (Phase 3 return)
- [ ] If missing ANY required tier field → **Needs Rework**
- [ ] If AI used in Tier 3 but poorly documented → **Needs Rework**
- [ ] If multiple vague statements ("some processing", "various corrections") → **Needs Rework**

APPROVAL DECISION MATRIX

Situation	Decision	Action
All tier requirements met	✅ **Approved**	Proceed to Phase 5
Missing/incomplete documentation, fixable in Phase 4	⚠️ **Needs Rework**	Return to Phase 4 with specific feedback
Fundamental quality issue (uncertainty < 3, corrupt data)	🔙 **Insufficient Plan**	Return to Phase 3 for re-acquisition
Wrong tier selected	🔙 **Insufficient Plan**	Return to Phase 1 to revise tier selection

---

FEEDBACK BEST PRACTICES

When Rejecting with "Needs Rework"
✅ **Do:**
- List specific missing items
- Reference field names explicitly
- Explain why each item is required
- Provide examples if helpful
- Estimate time to fix (helps planning)

❌ **Don't:**
- Say "incomplete" without specifics
- Reject entire submission for one small issue (if easily fixable)
- Use jargon without explanation
- Be discouraging (this is iterative!)

When Returning to Phase 3 ("Insufficient Plan")
✅ **Do:**
- Explain root cause (acquisition problem, not processing)
- Specify what needs improvement in Phase 3
- Provide technical recommendations
- Acknowledge this adds time but improves quality

❌ **Don't:**
- Return to Phase 3 for documentation issues only (use "Needs Rework" instead)
- Return without clear requirements
- Blame the team (it's a learning process)

EXAMPLES OF GOOD FEEDBACK

Example: Needs Rework
> "NEEDS REWORK: Processing parameters incomplete - missing surface reconstruction method and texture resolution (required for Tier 2). Manual intervention log improved but lacks operator names for 2 of 3 interventions. AI model needs validation metrics (F1-score or accuracy at minimum). Estimated fix time: 4-6 hours."

Example: Insufficient Plan
> "INSUFFICIENT PLAN: Uncertainty level (Geometry: 2, Color: 2) indicates fundamental capture quality issues - likely inadequate lighting. Recommend returning to Phase 3 with: 1) Cross-polarized lighting setup, 2) Color calibration target, 3) Increased photo count

(currently 85, recommend 200+). Current processing cannot compensate for acquisition problems."

Example: Approved
> "APPROVED: All Tier 3 documentation complete. Processing parameters comprehensive and reproducible. Manual intervention log exemplary (dates, operators, time spent documented). AI model fully documented including training data (2.3M tokens), validation metrics (F1: 89.3%), and availability. Ready for Phase 5: Ingest & Publication."

TIME ESTIMATES FOR REVIEW

- **Tier 1:** 10-15 minutes
- **Tier 2:** 20-30 minutes
- **Tier 3:** 45-60 minutes (AI documentation requires careful review)
- **Re-review after rework:** 15-20 minutes (focus on changed items)

COMMON PITFALLS TO WATCH FOR

1. **"Template Filling"** - Repeated text across projects without customization
2. **Version Amnesia** - Software mentioned without version
3. **AI Opacity** - "Used AI" without any details
4. **Intervention Invisibility** - Manual work done but not logged
5. **Uncertainty Inflation** - Rating "5" when "4" is more honest
6. **Parameter Vagueness** - "High quality" instead of specific settings

---

WHEN IN DOUBT

- **Ask questions** - Better to clarify than approve incomplete work
- **Request examples** - "Show me the manual intervention log format you used"
- **Check with domain expert** - Especially for AI models and uncertainty levels
- **Review past approved projects** - Maintain consistency
- **Consider project context** - Simple museum display vs. research publication have different standards

---

FINAL CHECKLIST BEFORE APPROVAL

- [ ] I have reviewed all required fields for the project's tier
- [ ] Cross-tier consistency checked
- [ ] If AI used, documentation is publication-ready
- [ ] If uncertainty < 3 in Tier 3, I've considered Phase 3 return
- [ ] My feedback is specific and actionable
- [ ] Decision (Approved/Needs Rework/Insufficient Plan) is clearly stated
- [ ] Comments explain the reasoning

Reviewer Signature: _______________
Date: _______________
Decision: Approved / Needs Rework / Insufficient Plan

Checklist Version: 1.0
Last Updated: 2026-01-16
Purpose: Standardize Phase 4 validation across reviewers

13.3.5 Phase 5: Ingest and Publication

Interpretation

Phase 5 Interpretation

Example 1-2: Simple
Straightforward publication. Tier 1 (web only), Tier 2 (+ DOI).

Example 3: Complex Rights (Tier 3, 2 reworks)
Attempt 1: "Public domain" - insufficient (site public ≠ model public)
Attempt 2: Rights fixed, but paradata missing (must exist BEFORE publication)
Attempt 3: ✅ Approved
- Rights stratified (site/photography/model separated)
- 47-page paradata report (comprehensive)
- 82GB preservation package (E57 + OBJ + metadata + paradata PDF)

Example 4: GDPR Crisis
Problem: 47 identifiable faces, apartment interiors, 23 license plates in 3D model
Cannot publish as-is - GDPR violation

Solution:
- AI face detection (YOLOv8) - 47 faces
- Manual blur (Gaussian 25px radius)
- 12 apartment windows redacted
- 23 license plates blurred
- Legal review confirms compliance
- **Redaction log = Tier 3 documentation**

Original archived (restricted access), redacted published (public)

Key lesson: Privacy > completeness. Document redactions thoroughly.

Version: 1.0 | 2026-01-17

Quick Guide

Phase 5: Ingest & Publication - Quick Guide

🎯 For Evaluators
Package: BPMN + Examples + Interpretation + Checklist + Guide
Key: Example 4 = GDPR redaction crisis (47 faces)

📋 Overview
Purpose: Publish & preserve digital assets
Duration: 1-5 days
Output: Published dataset + preservation package

🏊 Three-Tier Structure

Tier 1: Minimum Viable
1. Dataset Type (Preservation/Consultation)
2. Rights Statement (CC-BY/Public Domain)
3. Link to Original Source (inventory number)

Tier 2: Scientific
4. Dissemination Format (glTF/USDZ/OBJ)

5. Persistent Identifier (DOI/ARK)

Tier 3: Memory Twin
6. Paradata Report Link (comprehensive documentation)
7. Preservation Format (SIP - E57/OBJ/TIFF)
8. Reduction/Redaction Log (privacy/admin data removed)

📊 Examples

Ex	Object	Tier	Issue	Resolution
1	Vase	1	None	Simple web
2	Sculpture	2	None	DOI + repository
3	Castle	3	Rights unclear, no paradata	Stratified rights + 47-page report
4	Buildings	1→3	GDPR (47 faces)	AI detection + manual blur

****Lesson:**** Privacy compliance critical. Redaction = Tier 3 documentation.

Version: 1.0 | 2026-01-17

Validation

Phase 5 Validation Checklist

Tier 1
- [ ] Dataset type specified
- [ ] Rights statement (use RightsStatements.org URIs)
- [ ] Original source link

Tier 2
- [ ] Dissemination formats listed
- [ ] Persistent identifier (DOI/ARK)

Tier 3
- [ ] Paradata report exists (link provided)
- [ ] Preservation format specified (SIP structure)

- [ ] Redaction log (if applicable - privacy/admin data)

Red Flags
🚩 Vague rights ("public domain" without detail)
🚩 No paradata link for Tier 3
🚩 Privacy issues (faces/plates/interiors) not addressed
🚩 Rights confusion (site public ≠ model public)

GDPR Compliance
- [ ] Check for identifiable individuals
- [ ] Check for license plates
- [ ] Check for private property interiors
- [ ] If present: Require redaction OR restricted access

Version: 1.0 | 2026-01-17

13.3.6 End-to-End Process

Cross-phase decision guide

Cross-Phase Decision Guide

When to Return to Earlier Phase

Return to Phase 1 (Planning)
From Phase 2:
- Tier selection inappropriate for metadata requirements
- Budget insufficient for chosen tier

From Phase 3:
- Complexity grossly underestimated (safety issues)
- Equipment not available (need to revise method)

From Phase 4:
- Processing reveals tier too ambitious
- Quality unachievable with current tier

From Phase 5:

- Purpose unclear (affects publication strategy)

Return to Phase 2 (Metadata)
From Phase 5:
- Rights statement insufficient
- Aggregation requirements not met
- Standards conflict discovered

Supplemental Returns (Not Full Return)
Phase 3 Supplement:
- Add calibration targets (tier upgrade)
- Capture missing areas (coverage gaps)

Tier Evolution Decision Matrix

Situation	Action	Requirement
Stakeholder wants more quality	Consider 1→2 upgrade	Budget × 3-4, timeline +50%
Data better than expected	Consider 1→2 upgrade	Can supplement (not redo)
Complexity too high	Consider 3→2 downgrade	Accept reduced scope
Equipment failed	Hybrid same-tier	Methodology validation

Crisis Response Protocol

< 24 hours: Assess options
24-48 hours: Workshop with stakeholders + expert
48-72 hours: Decision + validation plan
Week 1-2: Test alternative before full commitment

Budget Reallocation Triggers

Acceptable:
- Cancel planned activity → invest in alternative (Journey 4: repair → turntable)
- Tier upgrade with new funds (Journey 2: conservation fund)
- Contingency use for re-capture (Journey 3: weather)

Unacceptable:
- Rob other phases (creates cascade failures)
- Exceed total budget without approval
- Hide overruns (transparency critical)

Version: 1.0 | 2026-01-17

Orchestration Guide

Orchestration Guide - How to Read Project Journeys

Purpose
This guide explains how to interpret the orchestration-level project journeys and use them for project management.

What Is Orchestration Level?

Phases 1-5: Task-level detail (what happens INSIDE each phase)
Orchestration: Project-level flow (how phases CONNECT and interact)

4 Project Archetypes

Journey 1: Happy Path (6 weeks)
Use as: Baseline for simple Tier 1 projects
Key trait: No iterations, smooth flow
PM Lesson: Achievable when scope matches tier

Journey 2: Tier Evolution (12 weeks)
Use as: Template for mid-project tier upgrades
Key trait: Tier 1→2 pivot at Week 5
PM Lesson: Data quality + budget enable evolution

Journey 3: Resilience (26 weeks)
Use as: Realistic Tier 3 complexity
Key trait: 13 submissions, multiple iterations
PM Lesson: Budget 40% extra time for Tier 3

Journey 4: Crisis Recovery (32 weeks)
Use as: Contingency planning reference
Key trait: Equipment failure → hybrid methodology
PM Lesson: Fast decision + validation = recovery

Decision Points Across Phases

When to Return to Earlier Phase

- **Phase 4→1:** Tier selection wrong (scope too ambitious)
- **Phase 3→1:** Complexity underestimated (safety issues)
- **Phase 5→2:** Rights unclear (metadata strategy insufficient)

When to Pivot Tier
- **Upgrade (1→2):** Opportunity emerges + budget available
- **Downgrade (3→2):** Complexity too high + timeline critical
- **Hybrid (3):** Equipment failure (Journey 4 example)

Timeline Estimation

Tier	Happy Path	With Iterations	With Crisis
1	6 weeks	8-10 weeks	12-15 weeks
2	10 weeks	14-18 weeks	20-25 weeks
3	20 weeks	26-32 weeks	35-40 weeks

Budget Contingency Usage

Journey 1 (Happy): 0% contingency used
Journey 2 (Evolution): Tier upgrade (4× budget)
Journey 3 (Resilience): 100% contingency used (€12K)
Journey 4 (Crisis): Under budget (€2K saved via pivot)

Red Flags vs Normal Iterations

🚩 Red Flags (Escalate)
- 3+ iterations on SAME issue (not learning)
- Budget overrun > 50% (scope too ambitious)
- Equipment dependency (single point of failure)
- Team demoralization (persistent failures)

✅ Normal Iterations (Expected)
- Tier 3: 2-3 reworks per phase (complexity)
- First attempt rework (learning)
- Cross-phase returns for tier evolution
- Adaptive pivots (external factors)

Using These Journeys

For Planning:
- Match your project to closest archetype
- Add 20-40% buffer based on complexity
- Identify potential crisis points

For Execution:
- Track submissions vs journey baseline
- Early warning if iterations exceed archetype
- Use decision trees for pivot points

For Retrospectives:
- Compare actual vs archetype journey
- Document deviations and reasons
- Build institutional knowledge

Version: 1.0 | 2026-01-17

Orchestration Project Journeys

Orchestration: Complete Project Journeys

Overview

This document presents 4 complete end-to-end projects through all 5 phases, showing the **orchestration level** - how phases interact, where projects return to earlier phases, and how real-world complexities emerge across the workflow.

Focus: Project management perspective, cross-phase decision making, tier evolution, failures and recovery.

Journey 1: Happy Path - Museum Exhibition (Tier 1)

Project: PROJ-2025-MUSEUM-15
Object: Collection of 12 decorative vases (19th century)
Institution: Municipal Museum
Initial Tier: 1 (Minimum Viable)
Duration: 6 weeks
Final Status: ✅ Success

Co-funded by the European Union

Timeline Overview

```
Week 1    Week 2    Week 3    Week 4    Week 5    Week 6
├──────────────┼──────────────┼──────────────┼──────────────┼
──────────────┼──────────────┤
| Phase 1 | Phase 2 | Phase 3 | Phase 4 | Phase 5 | LIVE    |
| ✅      | ✅      | ✅      | ✅      | ✅      | 🎉      |
└──────────────┴──────────────┴──────────────┴──────────────┴
──────────────┴──────────────┘
```

Week 1: Phase 1 - Planning & Complexity Assessment

Goal: Define project scope for online exhibition

Submission 1 (Day 3): ✅ Approved
Tier Selected: 1 (Minimum Viable)
Scope: 12 vases, basic photogrammetry, web 3D viewer for exhibition
Budget: €1,200 (€100 per object)
Purpose: Museum website enhancement (online exhibition "19th Century Ceramics")

Key Decisions:
- Tier 1 sufficient (web display only, no research component)
- Simple objects (ceramics, controlled environment)
- In-house photographer available
- No complexity challenges

Approval Time: 1 day (straightforward scope)

Week 2: Phase 2 - Metadata Strategy

Goal: Define how to describe objects

Submission 1 (Day 8): ✅ Approved
Standard: Dublin Core Simple
Language: Polish (institutional language)
Integration: Existing museum catalog database

Key Decisions:
- Dublin Core adequate for discovery (Tier 1)
- Single language (Polish audience)
- Metadata migrated from existing card catalog (already digital)

Approval Time: 1 day

Week 3: Phase 3 - Data Acquisition

Goal: Capture 12 vases

Submission 1 (Day 17): ✅ Approved
Method: Basic turntable photogrammetry
Equipment: Canon EOS R6, turntable
Operator: Museum photographer (Tomasz Kowalski)
Dataset: 12 objects × 60 images = 720 images (14GB)

Capture Schedule:
- Days 15-17: 4 objects/day
- Controlled studio environment
- Consistent lighting (LED softbox)

No Issues: Indoor, controlled conditions, simple objects

Approval Time: Same day (quality check passed)

Week 4: Phase 4 - Processing & Enrichment

Goal: Generate 3D models for web

Submission 1 (Day 22): ✅ Approved
Software: Meshroom (open-source, fits budget)
Optimization: Decimated to 100K triangles (web-friendly)
Output: 12 × glTF files for three.js web viewer

Processing:
- Automated batch processing (3 days)
- No manual intervention needed (simple geometry)
- Web optimization automatic

Approval Time: 1 day

Week 5: Phase 5 - Ingest & Publication

Goal: Launch online exhibition

Submission 1 (Day 29): ✅ Approved
Publication: Museum website (three.js viewer)
Rights: CC BY 4.0 (museum attribution)
Format: glTF only (web)

No Preservation Archive: Tier 1 = consultation copy only (acceptable for this use case)

Launch:
- Day 35: Online exhibition live
- Public reception: Positive (300 views first week)

Approval Time: 1 day

Week 6: Project Complete

Day 42: Exhibition live, project closed

Retrospective

What Went Right ✅
1. **Clear scope from start** - Tier 1 appropriate, no scope creep
2. **Simple objects** - Ceramics, controlled environment = predictable
3. **Realistic budget** - €1,200 adequate for Tier 1
4. **In-house expertise** - Photographer already on staff
5. **No iterations** - Every phase approved first attempt

Metrics
- **Total Duration:** 6 weeks (42 days)
- **Budget:** €1,200 (on budget)
- **Submissions:** 5 total (1 per phase, all approved)
- **Rework:** 0

- **Team Satisfaction:** High (smooth project)

Key Lesson
> **Tier 1 works beautifully when scope matches capability.** Don't over-engineer simple projects.

Journey 2: Tier Evolution - Renaissance Sculpture (Tier 1→2)

Project: PROJ-2024-SCULPT-89
Object: Wooden polychrome sculpture (Virgin Mary, 16th century)
Institution: National Heritage Institute
Initial Tier: 1 → **Evolved to Tier 2**
Duration: 12 weeks
Final Status: ✅ Success (with pivot)

Timeline Overview

```
Week 1-2  Week 3-4  Week 5   Week 6-8   Week 9-11  Week 12
├──────────────┼──────────────┼──────────────┼────────────────┼──────
──────────┼──────────────┤
│ Phase 1 │ Phase 3 │ PIVOT! │ Re-do  │ Phase 4 │ Phase 5 │
│ (T1)   │ (T1)   │ T1→T2  │ Ph 1-2-3 │ (T2)   │ (T2)   │
│ ✅     │ ✅     │ 🔄     │ ✅✅✅  │ ✅     │ ✅     │
└──────────────┴──────────────┴──────────────┴────────────────┴──────
──────────┴──────────────┘
```

Week 1-2: Phase 1 (Tier 1 - Initial Plan)

Submission 1 (Week 1): ✅ Approved (Tier 1)
Purpose: Basic documentation for museum website
Budget: €5,000
Tier: 1 (consultation copy only)

Plan:
- Simple web 3D viewer
- No research component

- Basic photogrammetry

Week 3-4: Phase 2 & 3 (Tier 1 Execution)

Phase 2 (Week 3): Dublin Core, approved ✅
Phase 3 (Week 4): Photogrammetry capture, 180 images

Capture Complete (Day 28)
- High-quality images (45MP camera available)
- Excellent detail captured
- Processing begins...

Week 5: THE PIVOT - Stakeholder Discovery

Day 33: Preliminary 3D Model Review

Stakeholders Present:
- Museum Director
- Conservation team
- Research curator (Renaissance art specialist)

Discovery:
> "This quality is extraordinary! We can see paint layer details, wood grain structure, conservation damage patterns. This could support the restoration study we've been planning!"

Researcher Request:
> "Can we upgrade this to research-grade? We need this for our conservation grant application (due in 6 weeks). We have additional budget (€15,000 available from conservation fund)."

Emergency Planning Workshop (Day 35)

Participants: Project manager, curator, conservator, budget office

Assessment:
1. **Current Status:** Phase 3 complete (Tier 1), processing started
2. **Data Quality:** Excellent (high-res capture already done)

3. **What's Missing for Tier 2:**
  - Quality standard documentation (FADGI)
  - Controlled vocabularies (Getty AAT)
  - Persistent identifier (DOI)
  - Versioning strategy
4. **Timeline:** 6 weeks available (tight but doable)
5. **Budget:** €15K additional approved

Decision: **TIER UPGRADE: 1 → 2**

Plan:
- Return to Phase 1: Update to Tier 2 planning
- Return to Phase 2: LIDO metadata (not Dublin Core)
- Phase 3: Already done (re-use existing captures + add calibration)
- Phase 4-5: Continue as Tier 2

Week 6: Return to Phase 1 (Tier 2 Revision)

Submission 2 (Day 38): ✅ Approved (Tier 2)
Purpose: Updated to include conservation research
Tier: 2 (Scientific)
Target Quality: FADGI 3-star
New Deliverables: Research publication support, conservation documentation

Budget Revised: €20,000 total (€5K original + €15K conservation fund)

Week 7: Return to Phase 2 (Tier 2 Metadata)

Submission 2 (Day 43): ✅ Approved
Standard: LIDO v1.1 (upgraded from Dublin Core)
Vocabularies: Getty AAT for materials/techniques
Versioning: Documented (card catalog 1987 → digital 2024)
Bilingual: English + Polish (for international research)

Week 8: Supplement Phase 3 (Calibration Addition)

Existing Data: 180 images (already captured) - REUSED ✅
New Capture (1 day): Calibration targets, ColorChecker

Submission 2 (Day 50): ✅ Approved

Smart Decision: Didn't re-capture everything, just added calibration to existing dataset

---

Week 9-11: Phase 4 (Tier 2 Processing)

Submission 1 (Day 70): ✅ Approved
Processing: Higher quality (5M polygons vs 500K for Tier 1)
Software: Metashape Professional (budget now allows)
Outputs:
- High-res mesh (research archive)
- Medium-res (web viewer)
- Orthophotos (conservation documentation)

Quality: FADGI 3-star achieved

Week 12: Phase 5 (Tier 2 Publication)

Submission 1 (Day 78): ✅ Approved
Publication:
- Institutional repository (with DOI)
- Museum website (web viewer)
- Europeana aggregation

Preservation Package:
- High-res OBJ + textures
- LIDO metadata
- Paradata report

DOI: 10.5281/zenodo.SCULPT-89

Week 12+: Impact

Conservation Grant Application (Week 14):
- Submitted with 3D documentation
- **Result:** €120,000 conservation grant AWARDED
- 3D model cited as key evidence for restoration need

ROI: €20K investment → €120K grant = 600% return!

Retrospective

What Enabled Success ✅
1. **High-quality initial capture** - Even Tier 1 was done well (reusable for Tier 2)
2. **Flexible planning** - Recognized opportunity mid-project
3. **Budget availability** - Conservation fund accessible
4. **Smart reuse** - Didn't re-capture, just supplemented
5. **Timeline realistic** - 6 weeks tight but achievable

The Pivot Point
Week 5, Day 33 - Stakeholder review of preliminary model
→ Discovery moment
→ Emergency workshop (2 days)
→ Tier upgrade decision

Metrics
- **Duration:** 12 weeks (vs 6 weeks original Tier 1 plan)
- **Budget:** €20K (vs €5K original)
- **Returns to Earlier Phases:** 3 (Phase 1, 2, 3 supplemental)
- **Submissions:** 8 total (some phases twice)
- **Final Impact:** €120K conservation grant

Key Lesson
> **Tier evolution is legitimate and valuable.** Mid-project tier upgrades work when: (1) Data quality supports it, (2) Budget available, (3) Timeline realistic, (4) Purpose clear. Don't rigidly stick to initial plan if better opportunity emerges.

Journey 3: Complex Resilience - Castle Documentation (Tier 3)

Project: PROJ-2025-CASTLE-007
Object: Medieval castle ruins (13th century, 45m × 30m site)
Institution: Institute of Archaeology
Tier: 3 (Memory Twin)
Duration: 26 weeks (6 months)
Final Status: ✅ Success (hard-earned)

Timeline Overview

```
Month 1    Month 2    Month 3    Month 4    Month 5    Month 6
├──────────────────────┼──────────────────────┼──────────────────────┼────────────────
────────┼──────────────────────┼──────────────────────┤
│ Phase 1  │ Phase 2  │ Phase 3  │ Phase 3  │ Phase 4  │ Phase 5  │
│ (3 tries)│ (2 tries)│ (attempt)│ (re-cap) │ (2 tries)│ (2 tries)│
│ ❌❌✅    │ ❌✅      │ ❌        │ ✅        │ ❌✅      │ ❌✅      │
└──────────────────────┴──────────────────────┴──────────────────────┴────────────────
────────┴──────────────────────┴──────────────────────┘

Total Submissions: 13
Total Approvals: 5 (one per phase, after iterations)
```

Month 1: Phase 1 - Planning (3 Attempts)

Attempt 1 (Week 1): ❌ Needs Rework
Problem: Complexity assessment too vague
Feedback: "40% moss coverage" (good), but missing safety assessment, no drone permits mentioned

Attempt 2 (Week 2): ❌ Needs Rework
Problem: Better detail, but still missing structural engineer sign-off for scaffolding, no contingency plan

Attempt 3 (Week 3): ✅ Approved
Complete: Safety docs, drone permits, weather contingency, budget breakdown (€85K with €12K contingency)

Delay Impact: 2 weeks (vs 3 days if done right first time)

Month 2: Phase 2 - Metadata (2 Attempts)

Attempt 1 (Week 5): ❌ Needs Rework
Problem: Multilingual strategy vague ("Polish and English" - who translates? QC process?)

Attempt 2 (Week 6): ✅ Approved
Complete: Translation workflow documented, LIDO→EDM crosswalk table, update policy, metadata rights (CC0)

Month 3: Phase 3 - Acquisition (Attempt 1 - Failed)

Capture Session (Weeks 9-11): Weather Window April

Week 9: Overcast (perfect) - 180 aerial images
Week 10: Overcast - 240 ground images
Week 11: Partly cloudy - 50 interior images

Submission 1 (Week 12): ❌ Needs Rework

Problems Discovered in Processing Test:
1. **Shadow issues** - Despite overcast, oak trees caused shadows 8-11am (east wall)
2. **Coverage gaps** - North tower interior insufficient (12 images, need 25+)
3. **ColorChecker obscured** - 3 targets hidden by vegetation regrowth (faster than expected in April)

Decision: Re-capture required (1 week)

Team Morale: Low (3 weeks in field "wasted")
Budget Impact: €8K spent on first capture, €6K for re-capture

---

Month 4: Phase 3 - Re-Acquisition (Attempt 2 - Success!)

Supplementary Capture (Weeks 14-15): May

Changes:
1. **Timing:** Afternoon only (1-4pm) - no tree shadows
2. **North tower:** Oblique drone angles (exterior shots into interior) - 40 additional images
3. **ColorChecker:** Replaced 3 obscured targets, higher placement (2m poles)
4. **Spiral staircase:** Added (was overlooked in first capture)

Submission 2 (Week 16): ✅ Approved

Final Dataset:
- 590 images total (420 original + 80 supplemental + 90 redundant)

- 76.8GB raw data
- Complete coverage, no gaps

Team Morale: Recovered (resilience celebrated)

Month 5: Phase 4 - Processing (2 Attempts)

Attempt 1 (Week 19): ❌ Needs Rework
Problem: AI model documentation incomplete (BERT mentioned but no training data, validation metrics)

Attempt 2 (Week 21): ✅ Approved
Complete: AI model (BERT) with training corpus (2.3M tokens Medieval Latin), F1-score 89.3%, ORCID identifiers for all creators

---

Month 6: Phase 5 - Publication (2 Attempts)

Attempt 1 (Week 23): ❌ Needs Rework
Problem: Rights statement unclear ("public domain" - site public ≠ model public)

Attempt 2 (Week 25): ✅ Approved
Complete: Rights stratified (site/photography/model separated), 47-page paradata report, 82GB preservation package

---

Week 26: Project Complete - Lessons Learned Workshop

Retrospective Meeting (Day 180)

Participants: Full project team + external reviewer

What Went Wrong:
1. **Phase 1 underestimation** - Didn't appreciate Tier 3 planning depth (2 weeks lost)
2. **Phase 3 field realities** - Weather/vegetation unpredictable (re-capture needed)
3. **Phase 4 AI documentation** - Team didn't understand "publication-ready" means full provenance

What Went Right:
1. **Team resilience** - Didn't give up after re-capture

2. **Budget contingency** - €12K buffer absorbed re-capture costs
3. **Iterative learning** - Each rejection improved next attempt
4. **Final quality** - Exemplary Tier 3 outcome

Cost-Benefit:
- **Investment:** €85K actual (€73K planned + €12K contingency used)
- **Deliverable:** Research-grade 3D documentation (published in "Medieval Archaeology")
- **Grant Impact:** €200K EU Horizon grant awarded citing this documentation
- **ROI:** 235% (€85K → €200K grant)

Metrics

Metric	Value
Duration	26 weeks (6 months)
Budget	€85K (100% of allocated budget)
Total Submissions	13
Total Approvals	5 (final approval each phase)
Rework Iterations	8 (across all phases)
Returns to Earlier Phases	0 (all within-phase iterations)
Field Days	25 (20 original + 5 re-capture)
Team Members	8 (archaeologists, photogrammetrists, metadata specialists)

Decision Tree: Key Crossroads

Crossroad 1 (Week 12): Re-capture Decision
Options:
- A) Accept coverage gaps (faster, cheaper)
- B) Re-capture (2 weeks delay, €6K cost)

Decision: B - Re-capture
Rationale: Tier 3 = complete documentation. Gaps unacceptable for research-grade.

Crossroad 2 (Week 19): AI Documentation
Options:
- A) Skip AI model detail (already mentioned)
- B) Fully document (2 days work)

Decision: B - Full documentation

Rationale: Reviewer correct - Tier 3 reproducibility requires full AI provenance

---

Key Lesson
> **Tier 3 is iterative by nature.** Budget 30-40% extra time for iterations. Resilience matters more than perfect execution. Every "failure" taught valuable lessons. Final quality justifies the struggle.

Journey 4: Failure & Recovery - Pottery Collection (Abandoned → Salvaged)

Project: PROJ-2024-POTTERY-198
Object: 347 Bronze Age pottery fragments
Institution: Archaeological Museum
Initial Tier: 3 (Memory Twin)
Duration: 32 weeks (8 months)
Final Status: ✅ Partial Success (salvaged)

Timeline Overview

```
Month 1-2    Month 3      Month 4      Month 5-6    Month 7-8
├────────────┼────────────┼────────────┼────────────┼────────────┤
│ Phase 1-2 │ Phase 3    │ CRISIS!    │ WORKSHOP  │ RECOVERY  │
│ (planning │ (scanner)  │ Equipment  │ Hybrid    │ Phase 3-5 │
│  Tier 3)  │ 180/347    │ failure    │ strategy  │ Complete  │
│ ✅ ✅      │ 52% done   │ 💥         │ 🔄         │ ✅         │
└────────────┴────────────┴────────────┴────────────┴────────────┘

PIVOT: Tier 3 single-method → Tier 3 hybrid methodology
```

Month 1-2: Phase 1-2 Planning (Ambitious Tier 3)

Phase 1 (Month 1): Tier 3 for 347 Objects
Plan: Every fragment documented at 0.1mm accuracy (structured light scanning)

Budget: €45,000
Timeline: 5 months
Team: Dr. Piotr Nowak (PI), 2 technicians

Approved: ✅ (ambitious but feasible with equipment)

Phase 2 (Month 2): Collection-Level Metadata
Smart Decision: Collection-level metadata + detailed for 20 exemplars
Approved: ✅

Month 3: Phase 3 - The Capture Begins

Week 9-11: Structured Light Scanning

Equipment: Artec Leo 3D scanner (€35K equipment, owned by museum)
Protocol: 23mm calibration sphere daily, 15 minutes/object

Progress:
- Week 9: 60 objects (great pace!)
- Week 10: 58 objects (still good)
- Week 11: 62 objects (on track!)

Day 78 (Mid-Week 11): 180/347 objects complete (52%)

Month 4: THE CRISIS

Day 78, 2:15 PM: Equipment Failure

Event: Artec Leo scanner GPU overheat error
Diagnosis: Hardware failure (out of warranty, purchased 2019)
Prognosis: Factory service required (8-12 weeks, €4,000 repair cost)

Emergency Assessment (Day 79)

Status:
- ✅ 180 objects scanned (120GB data, safe)
- ❌ 167 objects NOT scanned
- ❌ Scanner unusable
- ⏰ Project timeline: 3 months remaining

Options Analysis:

Option	Cost	Timeline	Quality	Viability
A) Wait for repair	€4K	+12 weeks	Same method	❌ Miss deadline
B) Rent Artec Leo	€10K	4 weeks	Same method	❌ Over budget
C) Buy new scanner	€35K	Immediate	Same method	❌ Way over budget
D) Alternative method	€2K	4 weeks	Different	✅ Possible
E) Abandon remaining	€0	0 weeks	Incomplete	❌ Unacceptable

Initial Team Reaction: Panic, frustration, blame ("should have bought warranty")

Week 14: Emergency Stakeholder Workshop

Participants:
- Dr. Piotr Nowak (PI, archaeologist)
- Dr. Elena Kowalczyk (Collections Director)
- Budget Office representative
- External consultant (Dr. Thomas Weber, 3D specialist)
- Conservation team

Workshop Objectives:
1. Assess methodological alternatives
2. Evaluate quality trade-offs
3. Determine budget feasibility
4. Make GO/NO-GO decision

Workshop Outcome: HYBRID METHODOLOGY

Decision: Pivot to photogrammetry for remaining 167 objects

Rationale:
1. **Quality comparable:** Close-range photogrammetry can achieve 0.08-0.12mm GSD (comparable to scanner's 0.1mm)
2. **Equipment available:** Museum has Nikon D850 (45MP) - already owned
3. **Cost:** Turntable (€1,500) + LED ring light (€500) = €2K total
4. **Timeline:** 3 weeks setup + capture (vs 12 weeks scanner repair)
5. **Validation:** Capture 20 exemplar objects with BOTH methods (cross-validation study)

Risk Mitigation:
- Cross-validation study proves methodological equivalence
- 20 objects × 2 methods = publishable comparison (turns problem into research opportunity!)
- If photogrammetry fails validation → still have 180 objects documented

Budget Reallocation:
- Cancel scanner repair: Save €4K
- Invest in turntable + lighting: €2K
- Net savings: €2K (under budget!)

Revised Plan: Hybrid Tier 3 - structured light (180) + photogrammetry (167)

Month 5-6: Recovery - Hybrid Methodology Execution

Week 15: Setup & Calibration
- Turntable installation
- Lighting rig setup
- Photogrammetry protocol testing
- 5 test objects captured (validation)

Week 16-19: Cross-Validation Study (20 Objects)

Protocol: 20 exemplar objects captured with BOTH methods
1. Structured light (Artec Leo - using existing scans from before failure)
2. Photogrammetry (new captures with turntable)

Analysis (Week 19):
- Geometric deviation: Mean 0.15mm, Max 0.4mm
- **Conclusion:** Methods comparable for pottery morphology (surface variation >> method difference)
- **Decision:** Proceed with photogrammetry for remaining 167 ✅

Week 20-24: Photogrammetry Capture (167 Objects)

Protocol:
- 72 images per object (360° turntable, 5° steps, 3 elevations)
- 100mm scale bar + coded targets
- X-Rite ColorChecker per session

Progress:
- 11 objects/day (slower than scanning, but acceptable)

- 15 working days total
- 12,024 images captured (890GB raw)

Completion: Day 168 (Week 24)

Month 7: Phase 4 - Hybrid Processing

Challenge: Two Processing Pipelines

Structured Light (180 objects):
- Artec Studio software
- Already processed (before scanner failure)
- 120GB OBJ files

Photogrammetry (167 objects):
- Metashape processing
- Batch alignment + reconstruction
- 3 weeks processing time

Submission 1 (Week 26): ✅ Approved

Key Documentation:
- Methodology note explaining equipment failure
- Cross-validation study results (20 objects, 0.15mm mean deviation)
- Justification for hybrid approach
- Both methods achieve FADGI 3-star quality

Reviewer Comment:
> "This is not a failure - this is adaptive research design. The cross-validation study is publishable. Equipment failure handled professionally. APPROVED."

Month 8: Phase 5 - Publication & Impact

Submission 1 (Week 30): ✅ Approved

Published Datasets:
1. **347 pottery fragments** (all Tier 3 quality):
  - 180 via structured light
  - 167 via photogrammetry
  - Collection-level metadata

2. **Methodological comparison study** (bonus publication):
   - "Structured Light vs Close-Range Photogrammetry for Bronze Age Pottery: A Cross-Validation Study"
   - 20 objects, both methods
   - Published in *Journal of Archaeological Science*

DOI: 10.5281/zenodo.POTTERY-198

Week 32: Retrospective - Turning Crisis into Opportunity

What Could Have Gone Wrong (But Didn't)

Scenario A: Abandon Project
- Lost 180 objects of work
- Budget wasted
- Demoralization

Scenario B: Wait for Repair
- 12-week delay
- Miss publication deadline
- Grant application impacted

What Actually Happened: Adaptive Success

Success Factors ✅
1. **Emergency workshop (48 hours)** - Fast decision-making
2. **Consultant input** - External expert validated alternative
3. **Cross-validation design** - Turned problem into research
4. **Budget flexibility** - Reallocation allowed (cancel repair → buy turntable)
5. **Team resilience** - Accepted hybrid approach

Unexpected Benefits
1. **Methodological publication** - Cross-validation study = bonus paper
2. **Equipment diversity** - Museum now has photogrammetry capability (new skill)
3. **Budget savings** - €2K under budget (scanner repair cancelled)
4. **Team learning** - Photogrammetry expertise gained

Metrics

| Metric | Value |

|--------|-------|
Duration	32 weeks (8 months)
Budget	€43K (€45K allocated, €2K under)
Crisis Point	Week 11 (52% complete, equipment failure)
Recovery Time	5 weeks (workshop → validation → proceed)
Final Objects	347/347 (100% complete, hybrid methods)
Publications	2 (dataset + methodological comparison)
Team Morale	High (crisis → innovation narrative)

Decision Tree: Crisis Response

Day 79: Equipment Failure Decision

```
Equipment Failed (180/347 done)
    │
    ├── Option A: Wait 12 weeks
    │   └── ❌ Miss deadline
    │
    ├── Option B: Rent scanner (€10K)
    │   └── ❌ Over budget
    │
    ├── Option C: Buy new (€35K)
    │   └── ❌ Way over budget
    │
    ├── Option D: Alternative method ✅ CHOSEN
    │   ├── Setup photogrammetry (€2K)
    │   ├── Cross-validate (20 objects both methods)
    │   ├── If validates → proceed (167 objects)
    │   └── If fails → hybrid publish (180 only)
    │
    └── Option E: Abandon
        └── ❌ Waste 180 objects of work
```

Key Lessons

1. Equipment Redundancy Planning

> **Tier 3 projects need backup plans.** Single equipment dependency = project risk. Future recommendation: Budget 5-10% for equipment contingency or identify alternative methods upfront.

2. Crisis as Innovation Opportunity
> **The cross-validation study turned failure into research contribution.** Published in *Journal of Archaeological Science*. Reviewers praised "methodological rigor in adverse circumstances."

3. Fast Decision-Making
> **Emergency workshop within 48 hours.** Avoided paralysis. External consultant input crucial (Dr. Weber validated photogrammetry viability).

4. Transparent Documentation
> **Methodology note in paradata explained equipment failure.** Reviewers appreciated honesty. "This is science - things go wrong. Document how you adapt."

5. Team Psychology
> **Reframing crisis as challenge.** Initial panic → workshop → validation → success. Team now tells this as "the pottery project where we innovated under pressure."

Epilogue: 1 Year Later

Impact Assessment:

Research:
- Primary dataset (347 pottery): Cited in 12 publications
- Methodological paper: 45 citations (unexpected impact!)
- Used in 3 PhD dissertations

Institutional:
- Museum now offers photogrammetry services (revenue: €15K/year)
- Equipment diversity = resilience for future projects
- Case study in project management training

Team:
- Dr. Piotr Nowak promoted (project leadership recognized)
- 2 technicians gained photogrammetry expertise (career development)
- External consultant (Dr. Weber) now regular collaborator

ROI:
- €43K investment

- 2 publications (one high-impact methodological)
- €15K/year new revenue stream
- Team capability expansion

> **Final verdict: Crisis handled professionally turned into institutional capability building and research innovation.**

Cross-Journey Comparison

Aspect	Journey 1 (Happy)	Journey 2 (Evolution)	Journey 3 (Resilience)	Journey 4 (Crisis)
Duration	6 weeks	12 weeks	26 weeks	32 weeks
Budget	€1.2K	€20K	€85K	€43K
Tier	1 (stable)	1→2 (evolved)	3 (stable)	3 (hybrid)
Submissions	5	8	13	7
Reworks	0	3 (tier upgrade)	8	1 (crisis pivot)
Crisis Points	0	1 (opportunity)	0 (iterations)	1 (equipment)
Returns	0	3 (Ph 1,2,3)	0	0 (pivot in Ph3)
Key Challenge	None	Tier evolution	Complexity execution	Equipment failure
Outcome	Simple success	High ROI	Research impact	Innovation
Lesson	Match tier to need	Flexibility pays	Resilience matters	Crisis → opportunity

---

Strategic Insights for Project Managers

When to Choose Each Tier

Tier 1: Choose When
- Purpose: Web visualization, general public access
- Budget: Limited (€1-5K)
- Timeline: Tight (< 2 months)
- Objects: Simple, controlled environment
- **Success Rate:** 90%+ if scope appropriate

Tier 2: Choose When
- Purpose: Research, conservation, scholarly publication
- Budget: Moderate (€5-50K)
- Timeline: Realistic (2-4 months)

- Need: Interoperability, standards compliance
- **Success Rate:** 75-80% with good planning

Tier 3: Choose When
- Purpose: Complex research, full provenance, reproducibility critical
- Budget: Substantial (€20-200K+)
- Timeline: Extended (4-8 months)
- Complexity: High (environmental, technical, methodological)
- **Success Rate:** 60-70% (higher iteration rate acceptable)

When to Evolve Tiers Mid-Project

Tier 1→2 Upgrade: Viable If
✅ High-quality initial capture (can supplement, not redo)
✅ Budget available (typically 3-4× original)
✅ Timeline allows (+50-100% time)
✅ Clear research justification emerges

Tier 2→3 Upgrade: Rarely Recommended
⚠️ Usually requires re-acquisition (Tier 3 needs calibration from start)
⚠️ Budget increase substantial (5-10× Tier 2)
⚠️ Better to restart with Tier 3 plan

Crisis Management Principles

Equipment Failure Response (< 72 hours)
1. **Hour 1-24:** Assess options (repair, rent, buy, alternative method, abandon)
2. **Hour 24-48:** Emergency workshop (stakeholders + external expert)
3. **Hour 48-72:** Decision + action plan
4. **Week 1-2:** Validation (test alternative before committing)

When to Pivot vs Persevere

Pivot When:
- Fundamental assumption wrong (equipment failed, tier inappropriate)
- Better opportunity emerges (funding, research impact)
- External constraints change (deadline, stakeholder needs)

Persevere When:
- Solvable through iteration (documentation gaps, minor rework)
- Learning curve (team improving with each attempt)
- Near completion (> 70% done)

Budget Contingency Guidelines

Tier	Base Budget	Recommended Contingency	Typical Use
1	100%	+10-15%	Minor rework
2	100%	+20-30%	Metadata iterations, QC
3	100%	+30-50%	Re-capture, equipment, iterations

Conclusion

These 4 journeys demonstrate that:

1. **Simple projects work beautifully** (Journey 1) when scope matches capability
2. **Mid-project evolution is valuable** (Journey 2) if data quality + budget support it
3. **Complexity requires resilience** (Journey 3) - iterations are normal, not failure
4. **Crises become innovations** (Journey 4) when teams respond adaptively

The orchestration level reveals what phase-level analysis cannot: how decisions ripple across the entire project lifecycle.

Document Version: 1.0
Date: 2026-01-17
Purpose: Orchestration-level view for project management

IMPULSE

IMmersive digitisation: uPcycling cULtural heritage towards new reviving StratEgies

Funded by the European Union