IMPULSE

IMmersive digitisation: uPcycling cULtural heritage towards new reviving StratEgies

Deliverable D3.4:

Guidelines on simplification of metadata based on open standards.

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**	30 Nov 2025
Version	6.0	**Submission Date**	27 Nov 2025
Related WP	WP3	**Document Reference**	D3.4
Related Task(s)	Tasks: 3.1.1; 3.1.3; 3.2.1; 3.2.3	**Document Type**	Report
Related Deliverable(s)	D17, D18, D19	**Dissemination Level**	Public
Lead Participant	KU Leuven	**Lead Author**	Theodora Rontzova (KUL)
Contributors	Bruno Vandermeulen (KUL)	**Reviewers**	Jan Tretschok (K8)
		Reviewers	Martin Gordon (FBKW)

Author(s)		
First Name	**Last Name**	**Partner**
Theodora	Rontzova	KU Leuven
Bruno	Vandermeulen	KU Leuven

Document History			
Version	**Date**	**Modified by**	**Modification reason**
1.0	15/09/2025	Theodora Rontzova (KUL)	Initial Document
2.0	30/09/2025	Bruno Vandermeulen (KUL)	Internal Review
2.0	20/10/2025	Theodora Rontzova (KUL)	Rework
3.0	27/10/2025	Bruno Vandermeulen (KUL)	Internal review
4.0	03/11/2025	Theodora Rontzova (KUL)	Ready for review
4.0	08/11/2025	Martin Gordon (FBKW)	Evaluation by assigned expert No. 1
4.0	20/11/2025	Jan Tretschok (K8)	Evaluation by assigned expert No. 2
5.0	24/11/2025	Theodora Rontzova (KUL), Bruno Vandermeulen (KUL)	Final draft, ready for final quality check
6.0	26-27/11/2025	Żaneta Żegleń (JU), Łukasz Pieczonka (JU)	Quality Control, editing to final version

Quality Control		
Role	**Who (Partner short name)**	**Approval Date**
Deliverable leader	Bruno Vandermeulen (KUL)	24 November 2025
Quality manager	Łukasz Pieczonka (JU)	27 November 2025
Project Coordinator	Żaneta Żegleń (JU)	26 November 2025

2 Executive Summary: Strategic Guidelines for Metadata Simplification in Digital Cultural Heritage and Immersive Environments

This deliverable presents the IMPULSE Consortium's strategic guidelines for simplifying metadata practices within the digital cultural heritage sector, with a specific focus on their application in immersive Multi-User Virtual Environments (MUVEs). This work is a primary outcome of Work Package 3 (WP3), established to address the growing need for efficient and easily applicable standardisation practices to facilitate data sharing across complex, interactive, virtual digital platforms. In an evolving landscape, simplification is positioned not as a reduction of quality, but as a strategic necessity for achieving the Findable, Accessible, Interoperable, and Reusable (FAIR) principles that underpin the IMPULSE project's overarching mission.

2.1 The Core Challenge: Navigating Metadata Fragmentation and Virtual Environments

The contemporary cultural heritage landscape is characterized by a "pluralism of standards," where the coexistence of schemas such as Dublin Core, CIDOC CRM, LIDO, MARC21 or others creates significant interoperability challenges, semantic mismatches, and inconsistent classification. This fragmentation places a considerable burden to aggregate data. Additionally, national and thematic networks that contribute to Europeana which must invest significant resources in harmonizing diverse datasets into the Europeana Data Model (EDM).

This long-standing issue is critically amplified in the context of emerging immersive platforms. A central finding of this report is the near-total vacuum of established metadata standards for MUVEs. Close to no metadata schemas or standards are present and currently implemented in these environments. This gap represents a critical barrier to the effective management, discovery, and preservation of immersive digital heritage experiences. It perpetuates the "cultural heritage paradox": extensive digitization efforts result in rich content that nonetheless remains underutilized due to systemic barriers in metadata and technical compatibility. The complex challenges already faced in aggregating 3D content for platforms like Europeana signal that extending existing models is insufficient for the even more complex, interactive "6D data" of MUVEs; a fundamentally new and more pragmatic approach is required.

2.2 Methodology: Combining Analysis with Practice-Based Reality

The recommendations presented in this deliverable are derived from a dual-pronged methodology that ensures they are both technically sound and operationally viable. The first part consists of an analytical framework. It involves a comprehensive review and comparative analysis of key metadata standards to assess their structural design, semantic depth, and suitability for multidimensional data (2D, 3D, 4D, and 6D). This is balanced by a second, practice-based orientation, which grounds the analysis in the day-to-day operational realities, workflows, and resource constraints of cultural heritage institutions of all sizes. This approach ensures that the proposed guidelines are not merely aspirational but are sustainable, adaptable, and directly address the practical barriers to implementation.

2.3 Updated Strategic Recommendations: A Pragmatic Framework for Interoperability in MUVEs

In response to the identified metadata vacuum in immersive environments, this report puts forth a set of updated, actionable recommendations designed to establish a baseline for interoperability without imposing prohibitive technical burdens.

- **Adopt Dublin Core as the Foundational Standard**: Given the absence of established norms in MUVEs, Dublin Core is proposed as a strategic starting point. Its lightweight structure, broad interoperability, and ease of implementation make it a pragmatic choice for creating a common, easily adoptable descriptive layer.
- **Embed a Persistent Identifier (PID) to the Original Metadata**: This is the most critical component of the strategy. The simplified Dublin Core record must contain a persistent, machine-readable link (such as a DOI or ARK) in the Identifier field that resolves to the full, rich metadatarecord at the source institution. This creates a sophisticated two-tiered system: a simple, universal discovery layer for broad interoperability in MUVEs, and a gateway to the deep, contextual source data. This approach solves the dilemma of simplification versus data loss, ensuring that provenance is maintained and that richer contextual information remains accessible.
- **Mandate a Minimal Set of Core Elements**: To ensure consistency, the IMPULSE Consortium has defined a minimal set of five mandatory Dublin Core fields. This shared schema serves as a foundational layer for consistent documentation across all partner institutions, facilitating effective integration and discovery of assets within the IMPULSE platform and beyond without excluding richer datasets to be uploaded.

Term Name	Definition	Comment	Mandatory
Contributor	An entity responsible for making contributions to the resource.	The contributing institution or entity.	x
Description	An account of the resource.	A brief, free-text summary of the cultural heritage object or asset.	x
Identifier	An unambiguous reference to the resource within a given context.	A Persistent Identifier (PID) linking to the full, rich metadata record at the source institution. This is a critical requirement.	x
Rights	Information about rights held in and over the resource.	A clear statement of rights (e.g., a Creative Commons license) governing the reuse of the digital asset.	x
Title	A name given to the resource.	The formal title or name of the object.	x

4. **Integrate AI-Enhanced Methodologies**: Institutions are encouraged to explore integrated workflows that combine traditional curatorial expertise with AI enhanced procedures to streamline the creation, enrichment, and management of metadata for MUVEs.

2.4 A FAIR and Inclusive Digital Heritage Ecosystem

The guidelines outlined in this deliverable advocate for a metadata ecosystem that is not only technically robust but also socially and operationally viable. The proposed simplification strategy acts as a bridge between the current realities of cultural institutions and the evolving demands of immersive digital environments. By establishing a simple, common baseline for MUVEs now, the cultural heritage sector has an opportunity to build an interoperable foundation from the ground up, avoiding the fragmentation of the past. This proactive framework empowers institutions of all sizes to participate meaningfully in the digital transformation, enhancing the quality and accessibility of their collections and contributing to a richer, more inclusive European cultural heritage landscape and the emerging Common European Data Space for Cultural Heritage.

3 Table of Contents

4 Abbreviations and Acronyms

Abbreviation / acronym	Description
2D	Two-dimensional
3D	Three-dimensional
4D	Four-dimensional
6D	Six-dimensional
API	Application Programming Interface
AR	Augmented Reality
ARCO	Architecture of Construction Objects
BuildM	Building Metadata Schema (from DURAARK project)
CARARE	Connecting Archaeology and Architecture in Europe
CCSI	Cultural and Creative Sector Industries
CIDOC CRM	CIDOC Conceptual Reference Model
CSV	Comma-Separated Values
DC	Dublin Core
E57m	Metadata schema for E57 3D point cloud data
EAD	Encoded Archival Description
EAF	Europeana Aggregators' Forum
EDM	Europeana Data Model
EFG	European Film Gateaway
EGI	European Grid Initiative
EMDaWG	Embedded Metadata Working Group
EUREKA 3D	EU Project for 3D Digitisation and Aggregation
EUscreen	Europeana Aggregator for Audiovisual Archives
Exif	Exchangeable Image File Format
FADGI	Federal Agencies Digital Guidelines Initiative
FAIR	Findable, Accessible, Interoperable, Reusable
glTF	GL Transmission Format
IIIF	International Image Interoperability Framework
IMCO	IMPULSE Community of Practice
IPR	Intellectual Property Rights
IPTC	International Press Telecommunications Council
ISO	International Organization for Standardization
JSON	JavaScript Object Notation
LIDO	Lightweight Information Describing Objects
LTI	Learning Tools Interoperability
MARC21	MAchine Readable Cataloguing (version 21)

METS	Metadata Encoding and Transmission Standard
MODS	Metadata Object Description Schema
MPEG-V	Media Context and Control
MR	Mixed Reality
MUSEU	Europeana Aggregator for Museums
OAIS	Open Archival Information System
OpenXR	OpenXR API Standard
OWL	Web Ontology Language
PID	Persistent Identifier
RDF	Resource Description Framework
SKOS	Simple Knowledge Organization System
SMART	Specific, Measurable, Achievable, Realistic, Timebound
SQL	Structured Query Language
STARC	Semantic Technologies for Archival Record Collections
VR	Virtual Reality
VRA Core	Visual Resources Association Core
WebXR	WebXR Device API
WP	Work Package
X3D	Extensible 3D
xAPI	Experience API
XML	eXtensible Markup Language
XMP	Extensible Metadata Platform

5 Background

5.1 IMPULSE Project

IMPULSE emerged out of the vision of a European immersive digitisation framework, driven by the forces of culture, creativity, storytelling, upcycled technology and safe, simplified standards. The project aims to synthesise innovative, multifaceted solutions and methodologies addressing the digitisation and accessibility processes of the collections that make up the field of digital cultural heritage.

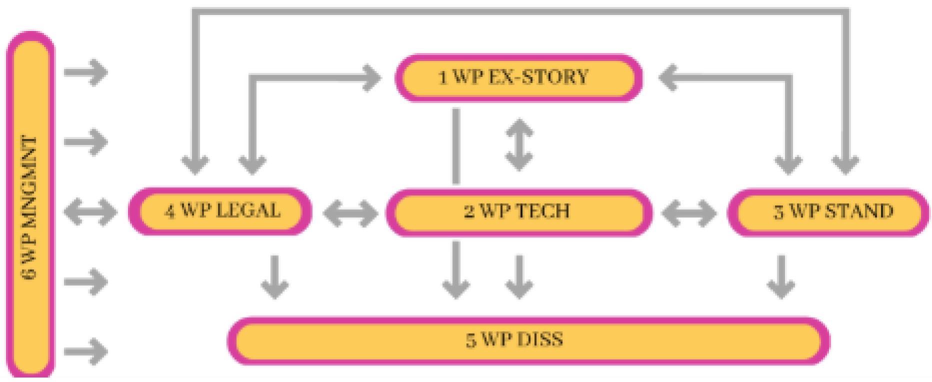
IMPULSE aims to address some of the most pressing gaps in the digitisation of European cultural heritage by building on existing knowledge, the capacity of its partners, and activities and networks. To achieve its stated intentions, the IMPULSE project has conceived a strategic plan which is divided 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.

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 standardisation protocols and revise 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) objectives, which entail:

- Solutions that will augment the quantity and range of cultural heritage objects displayed through VR/XR technologies. The now easily accessible collections shall function 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 that will enable the efficient (re)use of digitised 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 standardisation procedures and simplified strategies specifically targeted towards digitisation 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 utilisation of education, arts, and the CCSI.
- Legal and organisational frameworks with detailed evaluations of risks and barriers in fields such as the copyrights, 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 harmonisation.

- 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 digitisation.



To conclude, the overarching objective of IMPULSE is to create innovative and comprehensive solutions that foster the digitisation of cultural heritage in a standardised, 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 a growing need for efficient, easy to apply standardisation practices to facilitate data sharing across various platforms, and more specifically MUVEs.

Without standardisation, institutions may face challenges in sharing data. Some institutions may be reluctant to adopt certain standards as this comes with sometimes 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 customizes without the possibility for adaptation.

Well-established file formats and framework 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 professionals in the field and researchers, 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 the digital transition and digitisation, the very basics should not be underestimated. Correct and efficient standardisation practices in digitisation 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 standardisation practices data may become inaccessible and susceptible to loss or misinterpretation over time, compromising its value and significance. Most importantly, standardisation enables the creation of trustworthy data, ensuring it can be (re)used in the broadest manner.

More specifically, standardisation is important in the following areas:

- **Consistency**: Standardisation ensures that the data is captured, processed, and stored in a consistent manner, reducing the possibility of errors or unreliable datasets.
- **Interoperability**: Standardisation 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 understanding and interpreting the data.
- **Long-term preservation**: Standardised data is more likely to be preserved and be accessible in the future, ensuring its long-term availability.
- **Improved accuracy**: Standardisation provides clear guidelines and protocols for capturing, processing, and storing data, reducing the possibility of human error and improving accuracy.

6 Introduction: Guidelines on simplification of metadata based on open standards

Metadata is crucial for the management and sharing of digital heritage data. For 2D, 3D, and 4D data, there are well-established metadata schemas and practices with specific standards for different collection types (library, archive, museum) and specific standards such for the description of objects, object/collections (such as contextual, administrative and technical metadata, and paradata) for each phase of the data life cycle. However, there remains a significant gap in metadata standards, as well as in their implementation and searchability, particularly concerning immersive platforms and 6D data. In the scope of IMPULSE, we define 6D data as data used within MUVEs that integrate 2D, 3D, and/or 4D data and are reformatted into dedicated file formats to allow interaction in virtual spaces. The lack of standardised metadata for 6D data hampers the ability to make these data findable, accessible, interoperable, and reusable (therefore applying the FAIR principles) on novel immersive environments. This gap is predicted to present a critical barrier to the effective management and sharing of immersive digital heritage experiences as well as the accessibility and proper contextualisation of data, as the IMPULSE project is showcasing.

The current metadata landscape in the domain of cultural heritage is marked by significant fragmentation, both in terms of standards and implementation practices. Institutions across Europe and beyond rely on a wide array of metadata schemas, each developed with different priorities, structures, and levels of granularity. While these standards offer valuable frameworks for describing cultural heritage objects, their coexistence often leads to semantic mismatches, inconsistent classification, and interoperability challenges. This multitude of standards is further complicated by the emergence of new data types, particularly in immersive environments (e.g., 3D, 4D, and 6D data), which demand richer, more dynamic metadata that many existing schemas are ill-equipped to handle. Moreover, the adoption of metadata standards is uneven across institutions, with many smaller or under-resourced organizations relying minimal metadata or proprietary systems due to technical, financial, or capacity constraints. The lack of harmonization not only hampers data exchange and reuse but also undermines efforts to apply FAIR principles (Findable, Accessible, Interoperable, Reusable) across the sector. As a result, cultural heritage metadata remains siloed, difficult to aggregate, and often inaccessible to broader audiences or platforms such as Europeana. This is especially the case for aggregating data towards MUVEs. Addressing this fragmentation requires not only technical solutions but also a deep understanding of institutional workflows, capacities, and the practical realities of day-to-day digitization and documentation efforts.

In response to these challenges and particularly considering the near absence of established metadata standards within virtual platforms, the IMPULSE project has adopted Dublin Core as a strategic starting point. Its lightweight structure, broad interoperability, and ease of implementation make it suitable for MUVES, where technical resources and metadata embedding may be limited. Its simplicity allows collection holders, even those with limited technical expertise, to quickly adopt and apply without the need for specialized training or complex tools. Moreover, Dublin Core is highly adaptable and extensible, enabling the project to start with a basic metadata framework and gradually enrich it through refinements or mappings to more expressive models like CIDOC CRM or EDM, if needed. While it may not natively support the full complexity of multidimensional or immersive data, its flexibility ensures that it can serve as a solid foundation for metadata management, facilitating both immediate usability and long-term scalability. In this way, Dublin Core empowers IMPULSE to balance accessibility with long-lasting adaptability, making it an ideal choice for a dynamic and inclusive cultural heritage initiative. The IMPULSE Consortium has agreed on a minimum set of elementary Dublin Core fields as the base for the further development of the project: Contributor, Description, Identifier, Rights and Title.

6.1 Methodology

This deliverable adopts a dual-pronged methodology that combines a comprehensive overview and analysis of most commonly used metadata standards for the contextual description of objects with a practice-oriented perspective grounded in the operational realities of cultural institutions. Metadata creation, management and aggregation towards Europeana and other platforms is well-researched and, among others, part of previous deliverables of WP3 (Ongoing verification of the use of digital heritage objects within the emerging platforms) and aggregation strategies for the processes mentioned above are well-documented within the heritage field. As close to no metadata schemas or standards are present and currently implemented in MUVEs, our methodological focus is on the potential metadata aggregation towards emerging virtual environments based on the Dublin Core metadata schema.

6.1.1 Analytical Framework for Metadata Evaluation

This evaluation is conducted through comparative analysis of key standards such as Dublin Core, MARC21, CIDOC CRM, Spectrum, EDM, LIDO, and emerging schemas relevant to immersive platforms such as X3D, MPEG-V, glTF. Each standard is reviewed for its structural design, semantic depth, and extensibility, with particular attention to how it handles contextual metadata, paradata, and dynamic content.

The first phase of the methodology involves a systematic analysis of metadata types across several dimensions:

1. **Interoperability:** Evaluating the capacity of each metadata standard to integrate with other schemas and systems, including mapping potential and semantic compatibility.
2. **Reuse Potential**: Assessing how metadata facilitates the (re)use of digital heritage content in diverse contexts, including immersive environments, educational platforms, and creative applications.
3. **Suitability for Multidimensional Data**: Each metadata type is examined for its ability to support:
    - 2D data (e.g., images, texts)
    - 3D data (e.g., digital models, scans)
    - 4D data (e.g., time-based media, reconstructions)
    - 6D data (e.g., immersive, semantic-rich, and interactive environments)

6.1.2 Practice-Based Orientation

A defining feature of this methodology is its strong emphasis on the day-to-day practices of cultural institutions. Rather than proposing abstract or overly technical solutions, the guidelines are shaped by:

- **Institutional workflows:** We recognize the diversity of digitization practices, technical capacities, and resource constraints across museums, libraries, and archives.
- **Existing metadata standards in use:** We focus on the metadata standards already used within the IMPULSE consortium as well as in the broader field of (digital) cultural heritage. We acknowledge the technological and financial limitations of medium and small-scale cultural institutions; therefore, we advocate for the adaptive reuse of already existing standards.
- **Instances from practitioners**: To the degree that it is possible, we aim to incorporate insights from cultural heritage professionals regarding barriers to standard adoption, such as cost, technical complexity, and resistance to change.
- **Simplification strategies:** Within the framework of this deliverable's dual-pronged methodology, simplification strategies play a central role in addressing the complexities of metadata creation and aggregation in cultural heritage. Our approach is grounded in the principle of reducing unnecessary complexity while safeguarding the essential qualities that make metadata accessible, usable, and sustainable over time.

We prioritize lightweight, interoperable standards not as a replacement for existing models, but as a practical entry point for institutions operating under diverse constraints. These standards offer a balance between technical soundness and ease of implementation, making them particularly suitable for organizations with limited resources, technical capacity or metadata expertise. Simplification strategies are especially relevant in the context of multi-user virtual environments (MUVEs), where metadata practices are still emerging, and standardized schemas are virtually nonexistent. The absence of established norms in these platforms presents both a challenge and an opportunity: a challenge in terms of ensuring that metadata remains robust and interoperable across dynamic, immersive contexts and an opportunity to introduce adaptable frameworks that can evolve alongside technological innovation and allow for a higher degree of institutional flexibility and experimentation. By working with standards that are already familiar to many institutions, such as Dublin CORE for reasons analyzed below, we aim to facilitate smoother transitions into virtual spaces, enabling more effective sharing, discovery, and reuse of digitized cultural heritage assets. In this way, simplification becomes a bridge between current institutional realities and the evolving demands of digital and immersive environments.

6.2 Metadata standards within the IMPULSE Consortium

Metadata standards have occupied a central role within IMPULSE since the conception of the project. IMPULSE collections have been serving as a bounding link between the members of the Consortium and between the different work packages. D19 (Overview of new technologies (e.g. immersion platforms) in the field of data processing and sharing capabilities) of WP3 focused on defining a corpus of validated datasets, aiming to ensure the comprehensive and effective use of those collections. Within WP3, the partner institutions had the opportunity to present their collections and highlight key holdings. This process stood as a crucial introduction to the datasets and facilitated the compilation of a refined list of collections, and subsequently the development of a factsheet. The factsheet, with input from all work packages and all collections, captured essential details about the collections, including the metadata standards used within the collections and resulted in the internal Deliverable (D18) “Ground truth dataset for further usage within the project". In attachment 1 we present an overview of the metadata schemas and export file types used within each IMPLUSE institution and collection.

Within the IMPULSE Consortium the most frequently used metadata schema across partner institutions is Dublin Core, often in its extended form to accommodate domain-specific needs. While other schemas such as MARC21 and CIDOC-CRM are also present (particularly in libraries and specialized heritage collections) Dublin Core stands out for its lightweight structure, ease of implementation, and high adaptability across diverse collection types. This consistency within IMPULSE informed our decision to adopt Dublin Core as the primary schema for metadata harmonization. Starting from the specific context of IMPULSE, where institutions vary in size, technical capacity, and collection scope, Dublin Core offers a pragmatic solution that supports interoperability without imposing heavy technical requirements. More broadly, in the field of cultural heritage, Dublin Core continues to be a widely accepted standard among institutions of all scales due to its flexibility, cross-domain applicability, and support for incremental enhancement, making it an ideal foundation for inclusive and scalable metadata practices.

Most importantly, our decision to proceed with Dublin Core is informed by the previous research conducted within WP3. As “D3.1 Ongoing Verification of the use of digital heritage objects within emerging platforms” denotes that there is a vacuum in the metadata schemas used within MUVEs and the subsequent procedures of metadata management and simplification are still under development, we argue that Dublin Core is a safe and sound choice. It is a basic yet robust schema that most cultural professionals have at least some experiences with, therefore it was selected as our working metadata schema to be implemented in the IMPULSE virtual client. Additionally, the Dublin Core fields can also be partured or mapped within other metadata schemas.

7 Comprehensive Review of Common Metadata Standards Used within Cultural Heritage Institutions

7.1 Dublin Core

7.1.1 Introduction

Dublin Core is a metadata standard developed by the Dublin Core Metadata Initiative (DCMI) to facilitate the description and discovery of resources—both digital and physical—across diverse domains. It defines a set of 15 core elements (such as Title, Creator, Subject, Description, and Date) which provide essential information about a resource. The standard is designed to be easy to implement even by non-professionals, compatible across systems and platforms, and easy to expand with additional elements. Dublin Core is widely used in the field of cultural heritage, mainly in libraries, archives, educational repositories and supports file formats like XML, RDF and JSON-LD for machine-readable metadata.

7.1.2 Mandatory and Extended Elements

The Dublin Core standard is structured around two levels of metadata elements: simple (or elementary) Dublin Core, and Qualified (or Extended) Dublin Core. The core elements are 15 and provide the following descriptive information:

1. **Title** – Name of the resource.
2. **Creator** – Person or organization responsible for the content.
3. **Subject** – Keywords or topics covered.
4. **Description** – A summary or abstract of the resource.
5. **Publisher** – Entity making the resource available.
6. **Contributor** – Additional individuals or entities involved.
7. **Date** – Relevant date (e.g., creation, publication).
8. **Type** – Nature or genre of the resource (e.g., text, image).
9. **Format** – File format, physical medium, or dimensions.
10. **Identifier** – Unique reference (e.g., URL, ISBN).
11. **Source** – Related resource from which the current one is derived.
12. **Language** – Language of the content.
13. **Relation** – Links to related resources.
14. **Coverage** – Spatial or temporal scope.
15. **Rights** – Information about usage rights and access.

These elements are generic and flexible, making them suitable for a wide range of domains, including cultural heritage.

To support more precise metadata, Dublin Core also allows qualifiers and encoding schemes, among others:

dc:coverage.spatial – Geographic location relevant to the resource (e.g., archaeological site).

dc:coverage.temporal – Historical period or date range (e.g., 18th century).

dc:type.image – Specifies that the resource is a visual representation (e.g., painting, photograph).

dc:format.medium – Physical medium (e.g., oil on canvas).

dc:description.provenance – History of ownership or custody.

dc:relation.isPartOf – Indicates the resource is part of a larger collection or exhibit.

dc:subject.period – Historical or cultural period (e.g., Renaissance).

dc:rights.holder – Entity that holds copyright or ownership.

These extended elements are especially useful in museums, archives, and libraries, where detailed metadata enhances discovery, preservation, and scholarly research.

7.1.3 Interoperability

7.1.3.1 Dublin Core and LIDO Schema

While Dublin Core is simpler and more general, LIDO offers a more granular and event-centric model. Interoperability between the two is typically achieved through application profiles or crosswalks that map Dublin Core's general elements (e.g., dc:title, dc:creator) to LIDO's more specific structures (e.g., lido:objectIdentificationWrap, lido:eventWrap). This mapping often requires semantic enrichment, as LIDO supports multiple languages, hierarchical data, and detailed provenance, which Dublin Core does not natively accommodate

7.1.3.2 Dublin Core and Europeana Data Model (EDM)

The Europeana Data Model (EDM) was developed to integrate metadata from diverse cultural heritage institutions across Europe. It builds on Dublin Core but introduces semantic web principles and linked data structures. EDM distinguishes between the cultural heritage object (edm: ProvidedCHO), its digital representation (edm: WebResource) and the aggregation of metadata (ore: Aggregation), while Dublin Core elements can be reused within EDM (especially in the cultural heritage object class), but EDM adds complementary contextual classes. Mapping from Dublin Core to EDM is common and often involves refining or extending the original metadata to fit EM's richer structure.

7.1.3.3 Dublin Core and Spectrum

While Spectrum itself is not a metadata schema, it can be aligned with metadata standards like Dublin Core through crosswalks that link procedural data to descriptive elements. For example, a Spectrum procedure like "Object Entry" might be linked to dc:date, dc:source, or dc:rights. Interoperability among Dublin Core and Spectrum focuses more on harmonizing workflows and metadata outputs than direct element-to-element mapping.

7.1.3.4 Dublin Core and CIDOC CRM

Dublin Core and CIDOC CRM are interoperable through semantic mappings that allow the simple, general-purpose metadata of Dublin Core to be integrated into the rich, event-based ontology of CIDOC CRM. While Dublin Core provides basic descriptive elements like dc:title, dc:creator, and dc:date, CIDOC CRM offers detailed classes and relationships for modeling cultural heritage information, such as actors, events, places, and time spans. Mapping between the two enables cultural institutions to enrich Dublin Core records with deeper contextual meaning, supporting advanced data integration, reasoning, and linked data applications across museums, archives, and libraries.

7.1.4 Evaluation

7.1.4.1 2D data

Dublin Core is well-suited for describing 2D digital assets (that often comprise the majority of digitised assets belonging to heritage institutions) such as images, documents, and videos, thanks to its core elements like dc:title, dc:creator, dc:type, and dc:format.

7.1.4.2 3D data

For 3D objects, such as sculptures, architectural models, or digital 3D scans, Dublin Core can offer basic descriptive metadata. However, it lacks the ability to express spatial structure, geometry, or technical specifications of 3D models. While dc:type and dc:format can indicate that a resource is 3D, more detailed metadata often requires integration with specialized schemas (e.g., LIDO or CIDOC CRM) to capture the complexity of 3D representations.

7.1.4.3 4D data

Dublin Core's dc:date and dc:coverage elements can indicate temporal aspects, but it cannot model dynamic processes, event sequences, or temporal relationships. For meaningful representation of 4D data, Dublin Core must be extended or mapped to ontologies, which support event-based modeling.

7.1.4.4 6D data

6D data typically refers to highly complex datasets that combine spatial, temporal, contextual, and interactive dimensions—often used in engineering, simulation, or immersive heritage environments. Dublin Core is not designed to handle such multidimensionality. It lacks support for describing user interactions, sensor data, or multi-layered contextual relationships. Ontology-based models or domain-specific standards are necessary for adequate representation. When it comes to immersive environments, metadata must describe not only the content but also the user experience, navigation, interaction, and environmental context. While Dublin Core can provide basic metadata for the digital assets used in VR/AR (e.g., dc:title, dc:format), it is insufficient for modeling immersive narratives or spatial-temporal interactions. Integration with richer semantic frameworks is essential.

7.2 The Europeana Data Model (EDM)

7.2.1 Introduction

The principal aim of the Europeana Data Model has been to provide a qualitative change in the way that Europeana deals with the metadata gathered from aggregators and data providers.

It is important to first clarify that EDM adheres to the modelling principles underpinning the Semantic Web, meaning that there is no fixed schema that dictates just one or an optimal way to represent data. EDM can be seen as an "anchor" to which finer models can be attached and made (at least partly) interoperable at the semantic level. It does not require changes in the local approaches, although such changes are encouraged to the degree that they increase cross-domain usefulness, for example the use of publicly accessible vocabularies. In this sense, EDM has been an attempt to transcend the "individual" perspective of each community or institution that constitutes Europeana. It is not built on any particular community standard but rather adopts an open, cross-domain Semantic Web-based framework that can accommodate particular community standards.

EDM enables the representation and accessing of objects provided to Europeana via the packages of digital representations submitted by Europeana providers. In addition, EDM accommodates various description paradigms for the ingested objects and paves the way for enriching objects by connecting them to (networks of) semantically enriched resources. EDM does this while still allowing for different levels of granularity in the descriptions, using the possibilities of semantic mapping. This allows Europeana to retain compatibility with existing description approaches, including the simpler Europeana Semantic Elements (ESE) currently used for data submission at Europeana. It also provides support for ingesting the descriptive metadata submitted by various providers, possibly for the same object, and representing new information added by Europeana.

7.2.2 EDM Core and Extended Fields

The mandatory classes used to represent the essential structure of a cultural heritage object in EDM are the following:

- **edm:ProvidedCHO** – The actual cultural heritage object (e.g., a painting, book, or artifact).
- **edm:WebResource** – The digital representation of the object (e.g., an image or video file).
- **ore:Aggregation** – A container that links the ProvidedCHO and its WebResources, along with metadata like rights and provenance

Some other fields may not be mandatory, but they are highly recommended for richer data. Those usually are:

- **edm:Agent** – represents people or organizations related to the object (e.g., creator, publisher).
- **edm:Place** – provides geographical information (e.g., where the object was created or discovered).
- **edm:TimeSpan** – provides temporal information (e.g., date or period of creation).
- **skos:Concept** – for thematic or subject-related concepts (e.g., keywords, categories).
- **cc:License** – licensing information for digital resources

7.2.3 Interoperability

In principle, EDM is designed to be optimally interoperable with other metadata standards used in the cultural heritage domain. Below follows a brief analysis of the interoperability of EDM regarding its suitability for mapping to and from the dominant standards in the field.

7.2.3.1 EDM and Dublin Core

EDM reuses many Dublin Core properties (e.g., dc:title, dc:creator, dc:subject) for basic descriptive metadata. Mapping from Dublin Core to EDM is therefore relatively straightforward due to this reuse of properties, and EDM adds a semantic layer to the metadata.

However, when it comes to conversion from EDM to Dublin Core, EDM extends Dublin Core by introducing richer semantic relationships and contextual entities (e.g., edm:Agent, edm:Place) and the loss of contextual richness remains a high probability.

7.2.3.2 EDM and CIDOC CRM

EDM incorporates generalizations of CIDOC CRM concepts to support broader queries across heterogeneous datasets. A harmonized mapping between EDM and CIDOC CRM (specifically CRM-FRBRoo) has been developed, showing how EDM elements correspond to CIDOC CRM classes and properties- however some CIDOC CRM properties are not yet covered.

CIDOC CRM provides a more granular and event-centric model, which EDM simplifies for aggregation and interoperability purposes.

7.2.3.3 EDM and LIDO

LIDO is an XML-based schema designed for harvesting and sharing museum metadata, especially through OAI-PMH. LIDO and EDM share conceptual similarities, particularly in representing events, actors, and places. EDM is RDF-based, while LIDO is XML-based, so mapping requires transformation between data formats.

7.2.4 Evaluation of EDM

The Europeana Data Model (EDM) has been evolving to better accommodate complex digital cultural heritage representations, including 2D, 3D, 4D, and even 6D data.

7.2.4.1 2D data

The Europeana Data Model (EDM) is well-suited for representing 2D cultural heritage data such as images, texts, and audio files. It provides a clear structure for linking digital representations to physical objects, making it effective for standard digitization workflows. However, its support for interactive or layered 2D content is limited.

7.2.4.2 3D data

For 3D data, EDM has been extended to accommodate both reality-captured and digitally created models. It allows for multiple representations of the same object and maintains a clear distinction between the original and its digital versions. Despite these strengths, EDM does not yet offer standardized support for documenting the creation process of 3D models (paradata) or for embedding interactive 3D experiences.

7.2.4.3 4D data

In the case of 4D data, which involves temporal aspects such as historical reconstructions or time-based media, EDM can represent time spans and link reconstructions to original objects. Nevertheless, it cannot model dynamic changes within a single object over time, as its temporal metadata is static.

7.2.4.4 6D data

When it comes to 6D data, which integrates semantic, contextual, and multimodal dimensions, EDM excels in linking diverse types of information—such as people, places, and concepts—through a rich semantic framework. It supports complex relationships and aggregation of various media types. Still, it lacks native capabilities for handling real-time data, immersive environments, or sensor-based inputs, which are increasingly relevant in advanced digital heritage applications.

7.3 CIDOC CRM

7.3.1 General Introduction

The CIDOC Conceptual Reference Model (CRM) schema provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation. CIDOC CRM is intended to promote a shared understanding of cultural heritage information by providing a common and extensible semantic framework that any cultural heritage information can be mapped to. It has been intended to be a “common language” for domain experts and implementers to formulate requirements for information systems and to serve as a guide for good practice of conceptual modelling. In this way, it can provide the "semantic glue" needed to mediate between different sources of cultural heritage information, such as that published by museums, libraries and archives. For this purpose, it provides a structured framework to describe the concepts, relationships, and data used in cultural heritage documentation. Its event-centric documentation emphasizes the importance of documenting events to structure cultural metadata and their corresponding historical context. This approach contributes to the accurate representation of the life span of cultural objects by also involving human agents and interactions with the objects.

The CIDOC CRM standard is diverse and used to document a wide range of cultural heritage objects and data dimensions. The two-dimensional data types that can be described with this standard include but are not limited to flat representations like photographs, scans, paintings, drawings etc., with the possibility of incorporating their provenance and (historical) context.

The 3D data that is documented with CIDOC CRM includes three-dimensional models created by 3D scanning or modelling, with the additional possibility to encode metadata about production steps and digitization methods.

Time-based and volumetric data (four-dimensional data) may also be documented by CIDOC CRM, as their temporal aspects can be supported by the standard.

Higher data dimensions (for example 6D data for novel immersive environments), although less common, can also be supported by CIDOM CRM by including additional parameters or relationships between objects and events.

7.3.2 CIDOM CRM Elements

The CIDOC CRM standard is structured around a set of core and extended elements that comprise a comprehensive framework for documenting cultural heritage information.

The core elements are fundamental classes and properties that make up the backbone of the model. The following table lists 81 classes and the 160 Properties declared in CIDOC-CRM version 7.1.3. Below are listed in the classes and properties commonly associated with cultural heritage metadata.

CIDOC CRM Classes and Descriptions

- **E21 Person**: Individual involved in the creation, use, or documentation of heritage items
- **E39 Actor**: Person or group (e.g., institution) acting in a cultural heritage context
- **E22 Human-Made Object**: Tangible object created by humans (e.g., artifact, artwork)
- **E84 Information Carrier**: Physical medium bearing information (e.g., book, inscription)
- **E31 Document**: Information-bearing object (e.g., catalog, report)
- **E28 Conceptual Object**: Intangible creation (e.g., idea, text, plan)
- **E5 Event**: Occurrence involving heritage items (e.g., creation, discovery, exhibition)
- **E7 Activity**: Intentional human action (e.g., restoration, excavation)

- **E52 Timespan**: Temporal extent of an event or activity
- **E53 Place**: Geographical location relevant to heritage items or events

Some classes and properties may be extended. For example, the E55 (type) class comprises concepts denoted by terms from thesauri and controlled vocabularies used to characterize and classify instances of CIDOC CRM classes. It provides an interface to domain specific thesauri and ontologies that are represented as subclasses of E55, forming hierarchies that may be extended with additional properties.

7.3.3 Interoperability

CIDOC CRM is designed to facilitate interoperability between different metadata standards used in the documentation of cultural heritage. It provides a structured framework that can be mapped to various metadata standards, enabling the integration and exchange of information across different systems, which is crucial for fostering interoperability in the cultural heritage sector.

There exist established methodologies for mapping other metadata models to CIDOC CRM which tackle issues of heterogeneity and interoperability. CIDOC CRM supports the importing and exporting of data from and to other metadata standards through specific mappings and tools designed to facilitate these processes.

7.3.3.1 CIDOC CRM and Dublin Core

CIDOC CRM and Dublin Core are interoperable through harmonization efforts that map Dublin Core elements to CIDOC CRM classes. For example, Dublin Core's dc:creator maps to CIDOC CRM's E39 Actor, and dc:coverage can be expressed using CIDOC CRM's temporal and spatial entities like E52 Timespan and E53 Place. While Dublin Core is flat and general-purpose, CIDOC CRM offers a rich, event-based structure. Mapping from Dublin Core to CIDOC CRM is feasible and enhances semantic depth, but the reverse mapping often results in loss of granularity due to Dublin Core's simplicity.

7.3.3.2 CIDOC CRM and EDM

With EDM, CIDOC CRM serves as a conceptual backbone. EDM incorporates generalizations of CIDOC CRM concepts to support broad queries across aggregated datasets. A harmonized mapping exists between EDM and CRM-FRBRoo, showing how EDM's classes like edm:ProvidedCHO and edm:WebResource relate to CIDOC CRM's event and object-centric entities. However, EDM does not cover all CIDOC CRM properties, especially those dealing with complex temporal and spatial structuring.

7.3.3.3 CIDOC CRM to LIDO

Regarding LIDO, mappings to CIDOC CRM have been developed to align museum metadata with the ontology's event-based model. LIDO's structure, which emphasizes events, actors, and places, aligns well with CIDOC CRM's semantic framework. Mapping from LIDO to CIDOC CRM is conceptually coherent, but differences in data format (XML vs RDF) and granularity require careful transformation. The reverse mapping is also possible but may simplify or flatten the rich semantics of CIDOC CRM.

7.3.4 Evaluation

CIDOC CRM is not explicitly designed around dimensional data (2D, 3D, 4D, 6D) in the way that engineering or geospatial models are, but it can accommodate and interoperate with such data through its event-centric, extensible ontology and its official extensions.

7.3.4.1 2D Data

CIDOC CRM provides robust support for representing 2D data such as photographs, maps, and technical drawings, which are often used to document or interpret cultural heritage objects. The ontology allows these images to be linked to the objects or events they depict through properties like P138 represents, enabling semantic connections between representations and their referents. For example, a historical map (E38 Image) can be linked to the geographic region it depicts (E53 Place) and the time period it represents (E52 Time-Span). This modeling approach ensures that 2D representations are not treated as isolated artifacts but as integral parts of a broader semantic network, enhancing interoperability and contextual understanding.

7.3.4.2 3D Data

CIDOC CRM is equally well-suited to handle 3D data, both in terms of physical cultural heritage objects and their digital counterparts. Physical artefacts are modeled using E22 Man-Made Object, which can be richly described in terms of their creation (E12 Production), use, and modification over time. For digital 3D models-such as those created through photogrammetry or laser scanning—the ontology uses E73 Information Object in conjunction with the CRMdig extension, which provides detailed provenance for digital creation processes. This allows institutions to document not only the 3D model itself but also the methods, tools, and actors involved in its production. By linking digital models to their physical originals and the events surrounding their creation, CIDOC CRM supports a comprehensive and semantically rich representation of 3D data.

7.3.4.3 4D Data

CIDOC CRM is fundamentally a 4D ontology, designed to model entities and their relationships across both space and time. It excels at representing events (E5 Event, E7 Activity), periods (E4 Period), and the associated spatial (E53 Place) and temporal (E52 Time-Span) dimensions. This event-centric approach allows for detailed documentation of historical processes, such as the creation, acquisition, or restoration of an object, and the roles played by various actors (E39 Actor). For instance, a conservation campaign can be modeled as an E7 Activity that occurred at a specific place and time, involved certain people, and affected particular objects. This capacity to model dynamic, temporal phenomena make CIDOC CRM particularly powerful for historical and cultural heritage data, where understanding change over time is essential.

7.3.4.4 6D Data

While not explicitly labeled as a 6D model, CIDOC CRM—especially when extended with modules like CRMsci (scientific observation), CRMgeo (geospatial data), and CRMdig (digital provenance)—can effectively represent complex, multi-dimensional data that includes semantic and behavioral aspects. This includes not only where and when something happened, but also why, how, and with what implications. For example, a virtual exhibition that includes 3D models, interpretive narratives, user interaction logs, and scientific analyses can be modeled using CIDOC CRM and its extensions. The ontology supports the documentation of causal relationships, interpretive frameworks, and sequences of actions, enabling a holistic view of cultural heritage that encompasses physical, digital, intellectual, and experiential dimensions. This makes it suitable for advanced applications in digital humanities, virtual reality, and smart heritage systems.

7.4 LIDO

7.4.1 General Introduction

The LIDO (Lightweight Information Describing Objects) schema is a metadata standard designed to represent information about cultural heritage material objects. It is formally defined in the XML schema language and is used to deliver metadata for diverse online services, including collections, databases, and portals of aggregated resources. Its strength lies in its ability to support the typical range of descriptive information about objects of material culture. It can be used for all kinds of objects, e.g., art, cultural, technology, and natural science and supports multilingual portal environments.

The LIDO schema is the result of a substantial redesign and enhancement of the CDWA Lite and museumdat schemas based on recommendations of the CDWA Lite/museumdat Working Group, community feedback and further CIDOC-CRM analysis. It mainly builds on CDWA and includes additional concepts to meet SPECTRUM requirements.

7.4.2 The LIDO schema Elements

LIDO's structure is divided into several key areas, such as 1) descriptive metadata that includes information about the object such as its title, date, materials, techniques and related works, 2) administrative metadata that covers the information about records themselves, including identifiers, sources, and legal rights and 3) event information, as LIDO allows for detailed event- based descriptions, capturing the history and context of the object. The LIDO schema includes several core elements that are mandatory for generating a valid record, validating that the essential information about an object is captured.

Six content elements are declared mandatory in LIDO as a minimum requirement for a LIDO-compliant record. These elements are a subset of the core categories of CDWA and are considered necessary to unambiguously identify an object or work. The motivation for selecting only a small number of mandatory elements was to offer a low threshold for transforming existing data into LIDO. Having only a few restrictions also leaves room for adapting the LIDO schema to different requirements. The convenience of easy data transformation, however, must be balanced against the risk of accepting poor metadata that often leads to unsatisfactory search results. Therefore, some elements in addition to the minimum set are strongly recommended to be used in the object description.

The mandatory elements must be contained in a LIDO record. If any of these are missing, a validating XML processor will reject the LIDO record as invalid. Note, that this validation concerns the syntactical conformance to the LIDO schema only; it does not refer to the element contents, whether they contain data, and if this is semantically correct or not.

The following six content elements are mandatory, listed in the sequence of appearance in a LIDO record:

1. **Record Identifier** (<lido:lidoRecID>): the unique identifier for the metadata record. This mandatory element serves to distinguish an individual LIDO record from any other record that may occur in a database, data repository, or any other aggregation of machine-processable records. The LIDO Metadata Record Identifier is preferably composed of an identifier for the contributor and a record of identification in the (local) system of the contributor. It is not required to be persistent.

2. **Object/Work Type (**<lido:objectWorkType>**)**: This element describes the type of object or work being cataloged. The term originates from the CDWA Lite category Object/Work, emphasizing that it encompasses not only works of art but also human-made everyday objects and natural specimens. It captures the intrinsic or defining characteristics of an object; its "isness". A good descriptor should convey the essential features of the object to support accurate indexing and retrieval and always use the most specific term available from the controlled vocabulary to ensure precise search results. Following this principle (often referred to as Cutter's rule of Specific Entry) enhances the effectiveness of a well-structured knowledge organization system.
3. **Title Set (**<lido:titleSet>**)**: the title or name of the object. The Title Set element holds values for the appellation of the object or work, such as a title proper for a work or a name by which the object is known. A title is critical to always have a human-readable text referpιvy to an object and making it distinguishable from similar objects in search results. There may exist multiple titles in a given language, as well as titles in different languages. One title must be marked as the preferred one in each language, if there is more than one title; all other titles are regarded as alternative ones. It is strongly recommended to provide a descriptive, concise title that indicates the most important features to be recognized briefly. If no title or name is available, a descriptive one should be constructed based on the object/work type and further characteristics sufficient to select and distinguish the object in information retrieval.
4. **Record Identifier (**<lido:recordID>**)**: A unique identifier for the object within the source system. The Record Identifier element is a text string uniquely identifying the record in the contributor's database or other recordkeeping system. It serves as a reference for all communication with the originator concerning the contents of the metadata record.
5. **Record Type (**<lido:recordType>**)**: type of record (cataloguing, acquisition etc.). The Record Type element indicates the cataloging level selected for the record in question. It represents the logical tier of the <lido> object record, whether it refers to a single item, a part thereof, or a group of objects. Objects or works may be described at the following levels of granularity, as recommended in the LIDO Record Type Vocabulary.
6. **Record Source (**<lido:recordSource>**)**: the institution or database. The Record Source element holds identifying information on the source from which, or where the <lido> object record was created. The source is usually the repository, institution or person creating the record in question.

In addition to the mandatory core elements, LIDO supports a wide range of extended elements that provide more detailed information about each object. Such elements are not mandatory but *recommended*. The mandatory LIDO elements are sufficient to identify an object or work unambiguously, given that metadata elements are applied correctly. In most cases, however, providing the bare minimum of metadata will usually not be enough to enable good retrieval of results in terms of findability and object discovery.

Compared to LIDO, the CDWA standard defines some more elements as required, marked "core" in the CDWA Overview of Categories. These are, besides, information on the creator and creation, particularly metadata for classification and subject matter. These elements will be further described below.

1. **Classification (**<lido:classification>**):** information about the classification of the object, such as its category and genre. Classification assigns an object to one or more classes from a shared class scheme. Like <lido:objectWorkType>, the Classification element is used for grouping similar objects so that they can be retrieved in a single search operation. Unlike Object/WorkType which classifies the object at the most specific level suitable, the classification element aggregates objects based on broad categories.
2. **Measurements (**<lido:measurementsSet>**):** details about the dimensions, weight and other physical attributes. Measurements Set contains information about the dimensions of the object, comprising the measurement type, such as height or width, the corresponding unit, and the measured value.
3. **Materials/Techniques (**<lido:materialsTech>**):** used to create the object. Materials/Techniques contain information about the substances, such as the medium or support, and the techniques or implements, either incorporated in the object in focus, or used in the production or modification of the object.
4. **Event (**<lido:event>**):** detailed descriptions of events related to the object such as its creation, discovery or exhibition. Event contains information about occurrences associated with the object in some way. The element is meant to be used in the following contexts, to refer to - an event the object participated in or was present at, e.g., its production, modification, or provenance as a series of events.
5. **Related Work** (<lido:relatedWork>): information about related subjects of work. Related Work contains information about an object that is directly associated with the object in focus. However, there is no general answer to whether or not a work should be linked as associated. The decision will depend on how the benefits in retrieval are estimated. Will retrieving both objects at one go be meaningful to users? Or will the relation of the objects lead to a deluge of unwanted search results? These questions should be weighed when establishing an associative relation between works.
6. **Subject (**<lido:subject>**):** descriptions of the subject/ theme depicted in or associated with the object. Subject contains information about what is shown in an object or what is a theme of the work in focus. Indexing subject matter is strongly recommended, since it is a primary access point in retrieval, and users quite often perform topical searches. Subject may occur as depicted items, themes, or narrative content, and refers to abstract concepts and equally refers to abstract concepts and named entities.

LIDO "borrows" some definitions from other schemas, following the principle that useful data types and elements defined elsewhere should be reused instead of redefined. Such elements can be geographic locations (from the Open Geospatial Consortium), concepts (from SKOS), or individuals (from OWL). All schemas are identified by namespace prefixes which must be declared using the xmlns Attribute, preferably on the outermost element of a LIDO record.

7.4.3 Interoperability

The design of LIDO was guided by the principles of the reuse of existing standards, the use of proven technologies, adaptability and facilitating interoperability. As metadata is being moved, transformed and interconnected at an increasing rate, LIDO is called to facilitate and encourage harmonization with other standards.

7.4.3.1 LIDO and Dublin Core

Mapping LIDO to Dublin Core involves aligning LIDO's detailed descriptive elements with Dublin Core's more general elements. For instance:

- **Title** in LIDO maps to **dc:title** in Dublin Core.
- **Creator** in LIDO maps to **dc:creator** in Dublin Core.
- **Date** in LIDO maps to **dc:date** in Dublin Core

This mapping ensures that essential information about objects can be shared across systems using Dublin Core, facilitating interoperability and data exchange.

7.4.3.2 LIDO and MARC 21

Mapping LIDO to MARC 21 involves translating LIDO's XML-based structure into MARC's field-based format. Key mappings include:

- **Title** in LIDO maps to **245 field** in MARC 21.
- **Creato**r in LIDO maps to **100 field** in MARC 21.
- **Date** in LIDO maps to **260 field** in MARC 21

This mapping allows libraries and other institutions using MARC 21 to incorporate detailed object metadata from LIDO into their catalogues.

7.4.3.3 LIDO and CIDOC CRM

Mapping LIDO to CIDOC CRM involves converting LIDO's event-oriented descriptions into CIDOC CRM's semantic framework. Key mappings include:

- **Event** in LIDO maps to E5 Event in CIDOC CRM.
- **Object** in LIDO maps to E22 Man-Made Object in CIDOC CRM.
- **Actor** in LIDO maps to E39 Actor in CIDOC CRM

This mapping enables the integration of LIDO metadata into the Semantic Web, allowing for richer, interconnected data representations.

When importing or exporting LIDO metadata to the standards mentioned above, some tools that are typically involved are the transformation scripts that convert LIDO XML into any desirable target format, accompanied by validation and integration protocols. These processes ensure that metadata can be seamlessly shared and utilized across different systems.

7.4.4 Evaluation

LIDO is a versatile schema designed for describing objects of material culture and natural heritage. As with all standards, it possesses specific strengths and weaker features. LIDO supports comprehensive descriptive capabilities and a wide range of descriptive information such as event-based descriptions, which are crucial for documenting the history and context of cultural objects.

The schema is designed to facilitate data sharing and integration across different systems and platforms, making it suitable for multilingual and multi-institutional environments and is highly adaptable, allowing for customization and extension to meet specific needs. This flexibility is beneficial for institutions with diverse collections and varying documentation requirements. Moreover, the schema was developed through community feedback and collaboration to ensure that it meets the practical needs of users.

However, the detailed and comprehensive nature of LIDO can make it complex to implement, especially for smaller institutions with limited technical resources.

A closer cross-examination along with other renowned standards such as Dublin Core would reveal that LIDO offers more detailed and structured metadata compared to Dublin Core, which is simpler and more general. This makes LIDO more suitable for detailed documentation but potentially more challenging to implement. LIDO also provides a more flexible and event-oriented approach compared to MARC 21, which is primarily used for library materials. Ultimately, LIDO’s event-based descriptions align well with CIDOC CRM’s semantic framework, facilitating rich, interconnected data representations. However, mapping between LIDO and CIDOC CRM requires careful consideration of semantic relationships and analogies.

7.4.4.1 2D data

LIDO is well-suited for describing 2D objects due to its comprehensive descriptive capabilities. The Title and Creator, materials, measurements, or event information are all essential elements for identifying and attributing 2D objects. The same LIDO elements also support the detailed documentation of 3D objects as well as for time-based, four-dimensional data.

More complex data types such as 6D data extend beyond the spatial and temporal dimensions to include additional contextual information, such as environmental factors or user interactions. While LIDO is primarily designed for 2D, 3D, and 4D data, it can also be adapted to handle 6D data through extensions and customizations. LIDO's event-based structure proves useful in this context as well, can capture complex interactions and environmental contexts, making it suitable for documenting 6D data. It also supports linking to other objects or data points that are also inclusive of additional dimensions.

7.4.4.2 3D data

When it comes to 3D data, LIDO performs quite well in terms of metadata representation. It can describe the digitization process of 3D models, including the tools, actors, and events involved. It also allows institutions to link to external 3D files using the <lido:resourceRepresentation> element. However, LIDO does not natively support the encoding of 3D geometry or spatial data. It relies on external systems to store and render the actual 3D content, while LIDO provides the descriptive metadata.

7.4.4.3 4D data

For 4D data, which includes time-based media such as animations or video reconstructions, LIDO offers moderate support. It can represent temporal aspects through its event model, using elements like <lido:eventDate> and <lido:displayDate>. This allows for the documentation of performances, exhibitions, or conservation activities over time. However, LIDO does not include elements for encoding playback parameters such as frame rate or duration, which limits its ability to fully describe time-based media without relying on external metadata standards.

7.4.4.4 6D data

When considering 6D data, LIDO's capabilities are more limited. While it can describe semantic aspects using controlled vocabularies and authority files, and it can represent temporal changes through its event structure, it lacks the ability to model spatial coordinates, transformations, or semantic segmentation within 3D models. Furthermore, it is not designed for real-time or interactive applications such as augmented or virtual reality, which are often integral to 6D data use cases.

In summary, LIDO is a schema well-suited for descriptive metadata, particularly for cultural heritage contexts. It is well-suited for 3D metadata, moderately capable for 4D, and limited for 6D applications. For more advanced spatial, temporal, or semantic modeling, LIDO should be complemented with other standards such as CIDOC CRMdig, CityGML, or IFC, depending on the specific requirements of the project.

7.5 MARC21

7.5.1 Introduction

The original MAchine Readable Cataloguing (MARC) format was developed at the Library of Congress in 1965-1966. As the usefulness of the format kept gaining international recognition, national variants were subsequently developed and from the 1980s and onwards, the original format became known as US MARC. Later, USMARC and CANMARC were 'harmonized' in 1997 resulting in the format now named MARC 21. The British Library adopted MARC 21 as its cataloguing format in June 2004. MARC 21 is a widely used standard for encoding bibliographic data in machine-readable format, primarily used by libraries. It allows for the detailed cataloging of items such as books, DVDs and digital resources, facilitating the sharing of bibliographic information between different systems and institutions.

The bibliographic format contains data elements mainly for the following types of material: books (monographic two-dimensional textual material), serials (two-dimensional, periodical textual publications), maps (two- dimensional cartographic material), printed and manuscript- notated music (two- dimensional), sound recordings (time- based media), visual and mixed materials.

MARC 21 encodes several types of data across different formats. The main data types comprise of:

- **Bibliographic Data**: Information about books, journals, audiovisual materials, and other resources. This includes titles, authors, publication details, physical descriptions, and subject headings
- **Authority Data**: Standardized information about names, subjects, and titles to ensure consistency in cataloging
- **Holdings Data**: Details about the specific copies of items held by a library, including location, availability, and circulation status
- **Classification Data**: Information used to categorize items within a library's classification system
- **Community Information**: Data about events, services, and organizations within the community

MARC records are composed of three elements: the record structure, the content designation and the data content of the record. The record structure is an implementation of the standard for information interchange (ANSI/NISO Z39.2 and ISO 2709). Thus, MARC records have three main components:

Leader: Data elements holding information used in processing the record. The first element in any MARC record.

Directory: A series of entries that record the tag (or field label), length and starting location of every variable field in a record.

Variable fields: Each variable field is identified by a three-character numeric tag.

Regarding the data dimensions, MARC21 data is predominantly two-dimensional, consisting of structured records that encode bibliographic information in a linear format, using fields and subfields.

MARC 21 Core and Extended Elements

In MARC 21, fields can be categorized as **mandatory, mandatory if applicable, or optional.**

Mandatory Fields

These fields must be present in every bibliographic record:

Leader: Contains information for processing the record.

Variable Control fields

- 001 - Control Number: Unique identifier for the record.
- 005 - Date and Time of Latest Transaction: Timestamp for the latest update.
- 008 - Fixed-Length Data Elements: Contains coded data elements.

Mandatory if Applicable Fields

These fields must be included if the information is available and relevant:

- 100 - Main Entry-Personal Name: Used if there is a primary author.
- 245 - Title Statement: Includes the title and statement of responsibility.
- 260/264 - Publication, Distribution, etc.: Details about the publication.
- 300 - Physical Description: Information about the physical characteristics of the item.

Optional Fields

These fields can be included at the cataloger's discretion:

- 020 - ISBN: International Standard Book Number.
- 500 - General Note: Any additional notes about the item.
- 650 - Subject Added Entry-Topical Term: Subject headings.
- 700 - Added Entry-Personal Name: Additional authors or contributors.

7.5.2 Interoperability

Mapping metadata standards involves translating essential elements from one standard to another, enabling seamless integration of core qualities and information exchange. It is a process essential for metadata interoperability, the integration of diverse datasets, as well as for ensuring the necessary consistency in metadata descriptions, improving overall metadata quality.

The mapping process of course presents several challenges, such as the varying levels of complexity that different standards might have, the "untranslatability" of some elements that might lead to information loss, or the need for constant updates according to the changes in the original metadata standards.

A central component of the mapping process is the *crosswalks*, the specific mappings between standards that define how elements from one standard correspond to elements in another, thus facilitating the conversion of metadata from one standard to another. Once an institution identifies the elements of the source standards that correspond to the target standard, some basic conversion rules need to be established- for example, it should be clarified whether mappings will be done on a one-to-one basis or in a more complex manner such as linking multiple fields together or splitting them.

The mapping process can be implemented with software tools or manual processes that apply the crosswalks and covert metadata. (I can provide an example here?), with subsequent testing through validation tools.

7.5.2.1 Dublin Core and MARC21

Title

- MARC 21 Field: 245
- Conversion Rule: Direct mapping

Creator

- MARC 21 Field: 100
- Conversion Rule: Direct mapping

Subject

- MARC 21 Field: 650

- Conversion Rule: Direct mapping

Description

- MARC 21 Field: 520
- Conversion Rule: Direct mapping

Publisher

- MARC 21 Field: 260$b
- Conversion Rule: Direct mapping

Date

- MARC 21 Field: 260$c
- Conversion Rule: Direct mapping

7.5.2.2 MARC 21 and EDM

The interoperability between MARC 21 and the Europeana Data Model (EDM) is a structured process that enables libraries to contribute their metadata to Europeana's digital platform. MARC 21, a traditional library cataloging format, encodes bibliographic information in a standardized but relatively rigid structure. EDM, on the other hand, is designed for the semantic web and supports richer, more flexible metadata representations.

To bridge the gap between these models, Europeana has developed mapping guidelines that translate MARC 21 fields into EDM properties. This involves converting bibliographic elements like titles, authors, and publication dates into corresponding Dublin Core and other RDF-based elements used in EDM. The process also includes semantic enrichment, allowing MARC data to be linked to external vocabularies and resources, thereby enhancing its discoverability and contextual depth.

EDM's architecture supports the separation of physical and digital representations of objects, and it allows for the inclusion of contextual entities such as agents, places, and time periods. This makes it possible to preserve the integrity of the original MARC records while embedding them in a more interconnected and expressive data environment.

Europeana provides documentation and training resources to help institutions carry out this transformation, ensuring that library metadata can be seamlessly integrated into its broader cultural heritage ecosystem.

7.5.2.3 LIDO schema and MARC 21

Converting LIDO to MARC is not straightforward due to differences in structure, semantics, and domain focus. However, interoperability can be achieved through several mapping strategies and intermediate formats.Specific LIDO elements like the *title* or the *event* set can be translated to MARC fields such as 245 (title statement), 260/264 (the publication info) or 655 (Index term/genre). However, institutions should keep in mind that LIDO is an event-centric schema, and the support of additional properties often comes with significant simplification when converting to MARC.

Some institutions may use “intermediate” schemes to bridge similar gaps. Dublin Core can be considered as such due to its high interoperability potential- as it possesses a small and lightweight set of elements that can be easily applicable across domains and is widely adopted in the cultural heritage domain. However, the loss of some data granularity or the semantic mismatch when moving from an object-oriented schema to an event-centric one are challenges that have not been fully addressed.

7.5.2.4 CIDOC CRM and MARC 21

Converting CIDOC CRM to MARC 21 requires the bridging of two fundamentally different metadata paradigms. The conversion would require the flattening of several CIDOC CRM entities into MARC fields, with the subsequent loss of contextual details.

7.5.3 Evaluation

MARC 21 is a well-established metadata standard for encoding information in machine-readable form that is widely used in library institutions for cataloguing and exchanging data. It naturally presents different levels of suitability for more complex data types.

7.5.3.1 2D Data

MARC21 is highly suitable for 2D data, including conventional bibliographic materials such as books, journals, maps and other printed or digital materials that are two-dimensional and/or text-based. The standard provides comprehensive fields and subfields to capture detailed bibliographical information such as titles, authors, subjects, physical descriptions or publication data.

7.5.3.2 3D Data

When it comes to three-dimensional objects, MARC 21 is less directly applicable, as it does inherently support the detailed technical metadata required for 3D models. As this standard is structured to handle bibliographic information with a focus on 2d, textual data, it often lacks the granularity needed to describe some technical attributes of 3D models, such as geometric data, texture and material properties, or the spatial relationships between objects. Additionally, 3D data often involves interactive elements and dynamic properties, especially in contexts of virtual or augmented reality. MARC 21 currently does not support the metadata required to describe user interactions or animation within novel immersive environments.

While MARC 21 can be highly effective for traditional 2D bibliographic data, its application to 3D data is limited.

7.5.3.3 4D data

While it has been extended to support audiovisual and digital media, its structure remains fundamentally bibliographic and is not well-suited for the complexities of 4D data, which includes time-based media such as video, animation, or dynamic 3D reconstructions. MARC can capture basic information like duration and format, but it lacks the capacity to model internal temporal structures, dynamic changes, or interactive elements. Although enhancements like RDA integration have improved its flexibility, MARC still falls short for rich 4D metadata needs and is best used in conjunction with more specialized standards like MPEG-7 or PREMIS for comprehensive temporal and structural representation.

7.5.3.4 6D Data and Immersive Environments

Six-dimensional data types and environments present a specific complexity, to which the MARC 21 standard cannot respond adequately. Immersive environments specifically require metadata that can capture the spatial coordinates of 3D data for positioning purposes, temporal dynamics (for information that may change over time such as animations or user interactions), as well as the levels of interactivity within immersive platforms.

7.6 SPECTRUM

7.6.1 Introduction

The Spectrum standard is a widely recognized framework for museum collections management, developed by the Collections Trust in the UK and it has been adopted by museums worldwide. It serves as both a metadata standard and a procedural guide for managing museum collections. It is designed for museums of any size and any collection type and may also be useful to similar institutions with museum-like collections. Spectrum gives tried-and-tested advice on the things most museums do when managing their collections: daily activities, such as moving objects around and updating location records etc. Spectrum calls all these processes and activities as procedures and counts 21 of them. These can include the acquisition, cataloguing, conservation, loan and exhibition of items.

Each procedure has:

- A definition that tries to sum up the procedure in a single sentence.
- A fuller note on the scope of the procedure, which explains when to use it (and, at times, when to use a different procedure).
- The Spectrum standard. There are two parts to the standard: policy questions that need to be answered, and the minimum technical requirements that need to be met in the museum's written procedure. A suggested procedure could be a workflow diagram summarizing the suggested way of doing things, or a text version, which includes the information requirements for the procedure (see below). Where needed, the text versions of suggested procedures also include guidance notes.

The important thing to stress is that there is no one way to put any of these procedures into practice. Whether an institution uses paper-based systems, a computerised collections management system or – most likely – a mix of the two, it can adapt the suggested procedures to suit your needs. So long as the institution's way operational procedures meet the minimum requirements of the Spectrum standard, the documentation should be fundamentally sound. It is important to stress that there is no single way to implement those procedures, and they can be adapted to suit the needs of every individual cultural heritage institution.

7.6.2 Mandatory and Extended Elements

The Spectrum standard includes both mandatory and extended elements to facilitate the comprehensive documentation and management of museum collections.

Below we present the mandatory elements of SPECTRUM:

- Object Number: A unique identifier for each object.
- Object Name: The name or title of the object.
- Object Description: A detailed description of the object.
- Acquisition Date: The date the object was acquired by the museum.
- Condition: The current condition of the object.
- Location: The current location of the object within the museum

These elements provide additional information that can be recorded to enhance the documentation and management of collections. They are not mandatory but can be very useful for more detailed records. Some examples of extended elements include:

- Provenance: The history of ownership of the object.
- Exhibition History: Records of where and when the object has been exhibited.
- Conservation History: Details of any conservation work carried out on the object.
- Associated Documentation: References to related documents, such as research papers or photographs.
- Cultural Significance: Information about the cultural or historical significance of the object

These elements help museums maintain a rich and detailed record of their collections, supporting research, conservation, and public engagement.

7.6.3 Interoperability

The SPECTRUM metadata standard is widely used in the museum sector and is primarily designed to support the documentation of museum collections. Its interoperability with other metadata standards used in other cultural institutions is crucial for data exchange and long-tern preservation.

7.6.3.1 Spectrum to Dublin Core

Spectrum, as a collections management standard, provides structured procedures and units of information for museum activities. Dublin Core, being a general-purpose metadata schema, offers a lightweight and widely adopted vocabulary for describing digital resources. Interoperability between the two is feasible through basic metadata crosswalks, especially for descriptive elements like title, creator, and date. More advanced integration requires formal semantic alignment using RDF and the DCMI Abstract Model, allowing Spectrum data to be expressed in a way that supports linked data and validation through Description Set Profiles.

7.6.3.2 Spectrum to CIDOC CRM

CIDOC CRM, a formal ontology for cultural heritage information, offers the most comprehensive semantic interoperability with Spectrum. A detailed mapping has shown that the vast majority of Spectrum's units of information can be represented within CIDOC CRM's conceptual framework. This includes events, actors, objects, and relationships, allowing for rich, contextualized representations of museum data. Some Spectrum elements, particularly those related to legal and commercial aspects, may require extensions or clarifications within CIDOC CRM, but overall, the alignment is strong and well-documented.

7.6.3.3 Spectrum and EDM

The Europeana Data Model is designed for aggregating and linking cultural heritage metadata across institutions. Although Spectrum does not natively align with EDM, interoperability can be achieved by mapping Spectrum's concepts to EDM entities such as Provided Cultural Heritage Object (ProvidedCHO), WebResource, and Aggregation. This often involves intermediate schemas like CARARE, which bridge domain-specific metadata to EDM using linked data principles. Such mappings enable Spectrum-managed collections to be included in pan-European digital platforms like Europeana.

7.6.4 Evaluation

7.6.4.1 2D data

Spectrum is well- suited for two-dimensional data such as flat representations: drawings, photographs, maps etc. Specifically, Spectrum provides detailed procedures for cataloguing 2D objects, ensuring that all metadata is captured. It supports the proper documentation of 2D images and their associated metadata and is compatible with widely used standards for 2D, such as Dublin Core.

7.6.4.2 3D data

Spectrum procedures can also be adapted to include metadata specific to 3D objects, such as dimensions, materials, and techniques. As we discussed above, Spectrum can be mapped to standards like LIDO and CIDOC CRM, which support 3D data, facilitating the exchange of detailed 3D models. While Spectrum itself does not provide visualization tools, it can integrate with systems that offer 3D visualization capabilities.

7.6.4.3 4D data

4D data incorporates elements of time, such as changes over time or historical timelines. Spectrum can handle 4D data through its elements such as:

- Conservation History: Documenting changes and conservation efforts over time.
- Exhibition History: Recording when and where objects have been exhibited

7.6.4.4 6D data

Spectrum's suitability for 6D data mainly addresses issues such as the standard procedures for acquisition, conservation and deaccessioning support lifecycle management. Its ability to be mapped to standards like MARC 21 and CIDOC CRM somehow facilitates the management of complex data that involve multiple dimensions, albeit not to a maximum capacity.

Overall, Spectrum is highly adaptable and can be mapped to various other standards without compromising its interoperability and comprehensive documentation across different data dimensions. While Spectrum provides robust procedures for cataloging and managing collections, it may require integration with specialized systems for advanced visualization and sustainability tracking.

It is a versatile and effective standard for managing 2D, 3D, 4D, and (to a lesser degree) 6D data, especially when combined with other standards and systems.

7.7 Specialized/ Domain-Specific Metadata Standards

Some metadata schemas are not widely used but rather tailored to meet the unique needs of libraries, archives, museums, and architectural documentation projects. Below is an overview of more specialized metadata standards that support the encoding, transmission, and semantic enrichment of digital content across various disciplines.

7.7.1 METS (Metadata Encoding and Transmission Standard)

METS is a versatile XML-based framework that integrates descriptive metadata (e.g., MODS), administrative metadata (e.g., rights and provenance), and structural metadata to support the packaging of complex digital objects. It is widely used in digital repositories and long-term preservation workflows, enabling institutions to maintain the integrity and accessibility of digital collections over time. Use Case: Digital repositories, long-term preservation workflows.

7.7.2 MODS (Metadata Open Description Schema)

Designed for libraries and bibliographic records, MODS offers a richer alternative to Dublin Core while remaining simpler than MARC21. It uses XML elements such as <titleInfo>, <name>, <originInfo>, and <subject> to provide detailed descriptive metadata. MODS is often embedded within METS packages to enhance the discoverability of digital library objects.

7.7.3 EAD (Encoded Archival Description)

EAD is tailored for archives and special collections, providing a structured way to encode finding aids. Its hierarchical XML format reflects archival organization—from fonds to series, files, and items—and includes elements like <archdesc>, <dsc>, and <bioghist>.

EAD facilitates the contextual description of archival materials, preserving their provenance and original order.

7.7.4 VRA Core

VRA Core serves the visual culture and art history domains by describing works of art and their digital surrogates. It distinguishes between Work and Image records and includes metadata elements such as <title>, <agent>, <material>, <location>, and <date>. This schema is ideal for cataloging artworks, architecture, and visual media in museum and academic settings.

7.7.5 STARC (Semantic Technologies for Archival Record Collections)

STARC leverages semantic web technologies to model archival and cultural heritage metadata. It employs ontology-based descriptions and Linked Data principles to enhance interoperability and discoverability. STARC is commonly used in projects focused on digital preservation and semantic enrichment of archival content.

7.7.6 ARCO (Architecture of Construction Objects)

ARCO is a metadata model developed for documenting architectural heritage. It supports structured descriptions, semantic annotations, and 3D reconstructions of architectural objects. Often integrated with ontologies and semantic web technologies, ARCO facilitates the digital archiving and analysis of built heritage.

7.7.7 Buildm

Buildm is a metadata schema developed within the DURAARK project for preserving architectural data. It transforms input from IFC-SPF (Industry Foundation Classes - STEP Physical File) into a JSON-LD output format. Buildm captures building-specific metadata to support long-term digital preservation and reuse in architectural documentation.

7.7.8 E57m

Also, part of the DURAARK system, E57m is designed for 3D point cloud data derived from laser scans. It converts E57 input files into XML-based metadata, capturing details such as sensor specifications, scan positions, and data quality. E57m ensures that 3D imaging data is accurately documented and preserved for future use.

7.7.9 Croissant

Croissant is a metadata format built on schema.org to make datasets "machine learning-ready" by adding structured, machine-learning-specific information. It organizes metadata into four layers: general dataset details (like name, license, and creator), resource descriptions pointing to raw files, logical structure defining records and fields, and specific machine learning semantics such as splits, labels, and fairness attributes. This design enables interoperability, discoverability, and reproducibility across platforms and frameworks, allowing tools to seamlessly parse and load datasets for machine learning workflows.

7.8 The International Image Interoperability Framework (IIIF)

The International Image Interoperability Framework (IIIF) is a set of open standards designed to make it easier to deliver, share, and interact with high-quality digital images and audiovisual content online. It is widely used by libraries, museums, archives, and research institutions to provide consistent access to their digital collections. Within IMPULSE, IIIF is understood not as a descriptive standard but as a standard for sharing data, enabling interoperability and reusability across platforms and institutions.

While modern web browsers can display basic formats like .jpg and .mp4 at fixed sizes, their capabilities are limited beyond simple rendering. IIIF builds upon standard web technologies to enable far more advanced interactions with images and audiovisual content. For images, it provides functionalities such as deep zooming, side-by-side comparison, structured organization (e.g., maintaining page order in a digital book), and the addition of annotations. For audio and video, IIIF supports complex arrangements-such as multiple film reels forming a single movie-along with captions, transcripts, translations, and annotations.

By standardizing how these media objects behave, IIIF ensures consistent functionality across platforms and viewers. This interoperability allows content to be easily shared between institutions, thereby unifying collections and enhancing accessibility. In practical terms, IIIF defines several APIs that govern how digital content and metadata are delivered and presented. The Image API enables users to request specific regions, sizes, rotations, and qualities of an image; the Presentation API packages images and metadata for display; the Search API facilitates searching within annotations or transcriptions; the Authentication API manages access to restricted resources; the Content State API allows the sharing of specific views or configurations; and the Change Discovery API helps synchronize updates across systems. Together, these APIs form a cohesive framework that ensures reliable communication between servers and clients.

In this way, IIIF provides a robust foundation for data sharing and integration. Through its standardized APIs, institutions can publish digital images and audiovisual materials in interoperable formats that any IIIF-compliant viewer—such as Mirador or Universal Viewer—can display, regardless of where the content is hosted. Each digital object is described through a Manifest, a structured JSON file that includes metadata, sequence order, and media links, enabling the presentation of complex objects like books, manuscripts, or exhibitions. Because IIIF is open and extensible, it supports cross-institutional collaboration, allowing content from diverse sources to be seamlessly combined in digital exhibits, research tools, or websites. As a result, IIIF not only enhances accessibility but also promotes the portability, reusability, and longevity of digital cultural heritage.

7.9 Metadata Standards for Immersive Platforms

In the context of MUVEs, metadata standards are essential for structuring meaningful interactions, ensuring interoperability, and embedding contextual information that enhances user experience. Our approach, informed by D19 (embedding of metadata), emphasizes standards that support rich, dynamic, and sensory-aware environments.

7.9.1 X3D (Extensible 3D)

X3D is a foundational XML-based standard for representing 3D computer graphics. As the successor to VRML, it integrates well with other metadata standards and supports interactive 3D content across web and immersive platforms. X3D allows for the embedding of metadata related to sensory effects such as haptic feedback and ambient temperature, avatar behavior including gestures and emotional states, and virtual object properties like physics and interactivity. It also accommodates contextual metadata such as user preferences and environmental conditions. This makes it particularly useful for synchronizing real-world stimuli with virtual experiences, such as in VR theme parks or immersive training simulations.

7.9.2 MPEG-V (Media Context and Control)

MPEG-V (Media Context and Control) is designed to standardize interactions between virtual worlds and the real world. It defines metadata structures for sensory effects, avatar behavior, and virtual object interaction, making it ideal for immersive, multi-user platforms. MPEG-V bridges the gap between physical and digital environments by supporting geometric data, scene graph structures, animation metadata, and sensor routing. Its applications range from web-based 3D environments to educational simulations and collaborative design platforms.

In the domain of immersive learning, IMS Learning Tools Interoperability (LTI) combined with xAPI (Experience API) provides a robust framework for tracking learning experiences across platforms. These standards capture detailed metadata about user behavior, interactions, and outcomes, making them particularly valuable for gamified or educational metaverse platforms. Metadata elements include the actor (user), verb (action taken), object (target of the action), and context (device, location, session). This enables institutions to gather meaningful analytics from immersive learning environments such as VR classrooms and training simulations.

7.10 File Formats and Standards for Virtual and Immersive Platforms

While metadata interoperability is getting its much-deserved attention, the interoperability of file formats remains under-researched in projects and calls. The institutions involved in the IMPULSE project advocate for and apply an integrated approach.

D19 (Overview of new technologies in the field of data processing and sharing capabilities) conducted an in-depth analysis of file formats suitable for MUVEs 2D, 3D, and audiovisual file formats for virtual environment applications in cultural heritage digitization. It further integrated key outtakes from the other WP deliverables towards the joint effort to provide an integrated perspective on the challenges of using digitized cultural assets with complex file types in immersive contexts. For the evaluation of each file format's suitability, several (often competing) factors were taken into consideration: archival fidelity, real-time performance capabilities, interoperability across diverse software and hardware platforms, the richness of embedded or associated metadata, and considerations for long-term accessibility and preservation.

The systematic assessment has shown that file formats typically favored for archival purposes like TIFF, RAW camera files, high-polygon 3D models in PLY or OBJ formats, and lossless audio or video formats like FLAC and MOV, pose considerable challenges when integrated directly into MUVEs. These formats are often too large and require intensive processing, making them unsuitable for the real-time performance demands of immersive platforms. In contrast, formats specifically optimized for MUVEs such as JPEG for 2D images, glTF/GLB for 3D models, and MP4 for audiovisual content, are designed to support efficient rendering and reduced file sizes.

However, this optimization often comes at the cost of diminished data fidelity and limited metadata capacity. The cross-disciplinary work carried out within all IMPULSE WPs indicates that file format decisions are deeply linked to MUVE development challenges, the need for standardization, the digital heritage paradox and the potential of AI. The research conducted within the project on MUVEs reveals a consistent emphasis on visual fidelity. This often drives the selection of high-resolution and complex file formats, which, while visually rich, can significantly strain virtual environment performance unless optimized, typically through formats like glTF/GLB.

Additionally, the widespread use of custom-built MUVE platforms frequently results in the adoption of proprietary formats, which further complicates interoperability and limits long-term access to digital assets. These findings align with broader observations regarding the challenges of managing and sharing 6D data in immersive environments. Institutions often struggle to apply even well-established metadata and file format standards in these contexts, further underscoring the need for careful format selection. Therefore, efficient and interoperable file formats are critical for delivering immersive content across diverse platforms and devices. The following file formats not only support the transmission of 3D assets, but also embed metadata that enhances interactivity, personalization, and performance in MUVEs.

WebXR Device API, while not a metadata standard, provides essential access to VR and AR devices via the web. It enables metadata-rich interactions by supporting spatial tracking, user input, and device-specific extensions. Metadata embedded through WebXR includes materials and shaders, node hierarchies, animations, and morph targets, as well as extensions like KHR_lights_punctual and EXT_meshopt_compression. This API forms the backbone of immersive web applications, including game engines, AR apps, and lightweight 3D viewers.

OpenXR is a cross-platform API standard that facilitates the development of VR and AR applications across different hardware ecosystems. It supports metadata related to device capabilities, tracking data (head, hands, eyes), rendering layers, and interaction profiles. OpenXR promotes interoperability and scalability, making it a preferred choice for developers building XR applications for devices such as Meta Quest, HoloLens, and HTC Vive.

glTF (GL Transmission Format) is widely recognized for its efficiency in transmitting and loading 3D scenes and models. Often referred to as the "JPEG of 3D," glTF includes metadata for materials, animations, and spatial configurations. It supports user pose and motion tracking, input sources such as controllers and gestures, spatial reference frames, and session types like immersive VR and AR. This makes it ideal for real-time immersive web applications and collaborative XR experiences. Although it does not display built-in support for Dublin Core metadata, it allows custom metadata through specific mechanisms such as the use of extra fields to add the necessary metadata, and the construction of custom extensions.

Beyond 3D and XR-specific formats, traditional multimedia formats also play a role in immersive environments, especially when integrated into virtual experiences:

JPG with IPTC metadata is a widely used image format that supports embedded metadata for authorship, copyright, location, and descriptive tags. In immersive platforms, JPGs with IPTC can be used to enrich virtual galleries, training modules, or simulations with contextual information that are machine-readable and interoperable across systems.

MP3 and MP4 are standard formats for audio and video content, respectively. They support metadata such as title, artist, duration, and encoding parameters. In immersive environments, these formats are often used for background audio, voiceovers, and embedded video content, contributing to narrative depth and user engagement.

OGG/Vorbis, an open-source alternative to MP3, offers high-quality audio compression and supports metadata tagging. Its openness and efficiency make it suitable for immersive applications where licensing constraints or performance optimization are priorities, such as in open educational platforms or collaborative virtual spaces.

The importance of embedded metadata cannot be underestimated. Embedded metadata can "travel" with a digital object during its life cycle and often exists in synergy with metadata in an organization's database or in other information technology systems. Embedded metadata enables people in and outside of an organization to work more efficiently, provides valuable data to the systems that preserve digital content, and can assist in disaster recovery.

The Basic Guidelines for Minimal Descriptive Embedded Metadata in Digital Images, developed by EMDaWG, recommend a minimal core set of embedded metadata to supplement the standard fields used by institutions. Depending on the image file format, metadata (particularly IPTC data) can be stored within the file in various ways. Recent versions of image viewing and processing software increasingly save embedded IPTC data in XMP format. The required core set of metadata includes the **document title, copyright notice, source, and creator**, while the optional set may include elements such as the **date, description, keywords, credit/provider, job identifier, and headline.**

8 Aggregating data: the case of Europeana

8.1 Aggregators and Their Role Towards Europeana

Currently, there are no uniform platforms or aggregators that support virtual environments. The aggregation of data towards Europeana could function as an exemplary workflow for MUVEs, highlighting the good practices that can be adopted but also the shortcomings and challenges encountered, and how these can be avoided or mitigated. When it comes to metadata aggregation, we also propose Dublin Core due to the lack of other metadata schemas or standards in the field.

Aggregators play a crucial and foundational role within the European cultural heritage ecosystem. They function as essential intermediaries, bridging the gap between individual cultural heritage institutions and the large-scale Europeana platform. These aggregators are typically organizations such as libraries, archives, museums, or ministries of culture. Their primary task is to collect, standardize, and enrich digital cultural heritage data from institutions operating within their specific domain. Once processed, they share this harmonized data with Europeana, ensuring it fully adheres to the platform's rigorous quality and interoperability standards.

This intermediary function has led national and thematic aggregators to be viewed as a form of "middleware." This analogy is particularly apt when considering the broader context of Europeana itself, which functions as a multi-user virtual platform. Europeana serves a highly diverse set of users-including researchers, developers, educators, cultural professionals, students, and the general public-while bringing together thousands of contributors and offering services for discovery and search. (It is noted, however, that it lacks interactivity common to most commercial virtual platforms.)

Within this ecosystem, aggregators are tasked with assisting the participation of numerous users and institutions from their respective countries or domains. They bear the responsibility of ensuring their contributors are technically and semantically aligned with all of Europeana's standards.

8.1.1 The Technical and Functional Role of Aggregators

In technical and functional terms, aggregators are intermediary layers that perform several critical "middleware" functions:

- Data Mediation: This is perhaps the most critical function. Aggregators are responsible for the transformation and mapping of various metadata schemas-which differ from institution to institution-into the standardized Europeana Data Model (EDM). This alignment promotes consistency and is essential for interoperability.

- Quality Control: They perform quality checks before the metadata ever reaches Europeana, ensuring a higher standard of data across the platform.
- Rights Management: Aggregators play a key role in ensuring compliance with licensing and rights standards.
- Technical Integration: They provide the technical infrastructure for data flow, often offering APIs, OAI-PMH harvest endpoints, and other services to facilitate the smooth transfer of data from the institution to the aggregator, and onward to Europeana.

This aggregation function is not just a convenience; it is essential for the entire Europeana ecosystem. By handling these tasks, aggregators significantly alleviate the workload on Europeana, which would otherwise be burdened with managing thousands of individual, heterogeneous data sources.

Beyond these core technical duties, aggregators make a rich and essential contribution to the network. They actively foster continuous learning by equipping their partner institutions with vital training, specialized tools, and best practices for digitization, data management, and the aggregation process itself. Furthermore, national aggregators serve as national representatives, ensuring that their country's digital cultural heritage is accurately and meaningfully reflected at the European level.

This entire infrastructure is supported by a collaborative network, the Europeana Aggregators' Forum (EAF). All Europeana aggregators are members, and the forum allows national, regional, domain, and thematic aggregators to exchange knowledge and best practices, further supporting data sharing with Europeana.

8.1.2 National Aggregators

National and regional aggregators are defined by their geographic scope, working with contributors situated within a specific country or region. The text provided outlines several key examples for partners within the IMPULSE consortium:

Poland: Digital Libraries Federation (FBC) FBC is the only national cross-domain aggregator in Poland. It works with over 130 Polish data providers, including digital libraries, museums, and institutional repositories. A majority of these providers are regional digital libraries, which themselves host data for multiple institutions within their region. FBC is one of the largest aggregators for Europeana, providing over 2.1 million records. Its services extend beyond aggregation to include support on technical, organizational, and legal matters. FBC also handles data conversion (e.g., from CSV to XML), mapping to EDM in cooperation with the provider, basic enrichments, and preview generation. Technically, FBC strongly recommends an OAI-PMH interface but accepts CSV files as a minimum requirement and requires providers to have legal entity status.

Belgium: Erfgoedplus.be Starting as an initiative of two provinces in the Flemish region, Erfgoedplus.be has been managed by the Flemish ministry for culture since 2018. While open to all heritage, its main focus is supporting local, often non-professional, heritage holders in registering and sharing their knowledge according to professional standards, including semantic web technology. Its services include training, data analysis, mapping, enrichment, and a web-based application for registering small collections. Technical requirements include metadata in XML, a minimum standard data model, consistent use of formats and thesauri, and preferably accompanying images or other media.

Italy: CulturaItalia Managed by the ICCU (a branch of the Italian Ministry for Cultural Heritage), CulturaItalia is the national aggregator for Italy and has been active since 2008. It provides access to over 3.4 million digital objects from more than 600 Italian institutions and is interoperable with other thematic and regional portals. CulturaItalia offers extensive services, including legal advice on rights statements (Creative Commons, IPR), data analysis, mapping, data storage for museums, EDM conversion, and training courses. To work with CulturaItalia, providers must use the OAI-PMH protocol or upload data in CSV format, ensure digital media is available via a direct link, and have all content clearly rights-labeled.

Greece: SearchCulture.gr Developed in 2012 by the National Documentation Centre (EKT), SearchCulture.gr is Greece's national portal for aggregating digital cultural heritage and contemporary artistic production. It serves as a unified access point to a massive range of content, from archaeological monuments to audiovisual records. It collaborates with a wide array of institutions, from the National Gallery to local archives. Its services focus on data quality, offering guidelines and validation tools, semantic enrichment, publication as Linked Data, secure metadata backup, and training. It is flexible in its technical requirements, able to harvest metadata via OAI-PMH in seven different formats, including EDM, ESE, OAI-DC, and MODS.

Malta: Heritage Malta (HM) HM is the national agency for museums and cultural heritage in Malta, responsible for over 90 heritage sites. As an aggregator, HM contributes digitized assets from the national collection, including 2D, 3D, and audiovisual materials. It supports its partners with metadata standards and mapping, technical infrastructure, digitization services (both 2D and 3D), and guidance on Open Access and FAIR principles. HM requires metadata in XML or JSON (via direct export or OAI-PMH), accepts EDM and Dublin Core schemas, and prefers digital media to be online using IIIF-compatible services, oEmbed, or Sketchfab.

Germany: Deutsche Digitale Bibliothek (DDB) The DDB aims to offer unrestricted access to Germany's cultural and scientific heritage. An accredited aggregator since 2012, the DDB advises its providers on metadata formats, mapping, data quality, legal matters, and data delivery. It also offers a free tool, DDBstudio, for data partners to curate their collections. DDB is highly flexible, accepting a wide range of metadata formats

(METS/MODS, EAD, LIDO, EDM, etc.). Key requirements are that digital objects must be accessible online with a stable URL, have a stable ID, and (for images) have a resolution of at least 800x600 pixels.

8.1.3 Domain and Thematic Aggregators

In contrast to national aggregators, domain and thematic aggregators define their scope by a specific cultural sector (like museums) or a topic (like fashion). A key difference is that they typically work with contributors from multiple European countries.

CARARE Network (Archaeology and Architecture) CARARE aims to advance professional practice and appreciation for digital archaeological and architectural heritage. It provides extensive advice on metadata standards, data quality, mapping, and conversion to CARARE and EDM formats. It also offers specialized guidance on geographic data (e.g., converting national coordinates to WGS 84), subject indexing, temporal data, Linked Data, copyright, and 3D content. CARARE has developed its own CARARE metadata schema, a harvesting schema designed to capture detailed information about heritage assets (monuments, buildings, artefacts), digital resources, collections, and activities (like archaeological excavations) . This schema is based on standards like MIDAS Heritage and CIDOC CRM, and is compatible with LIDO and EDM. The schema's mapping process is critical, enabling the migration of data from native schemas, the aggregation of records from various standards, and the final transformation into EDM for Europeana.

PHOTOCONSORTIUM (Photography) This is Europeana's domain aggregator specializing in photographic content. It has made over 500,000 photographs accessible in Europeana from around 50 partners, ranging from prestigious public institutions to private collectors. This high-quality data is also featured in a dedicated thematic collection, "Europeana Photography," which PHOTOCONSORTIUM curates with virtual exhibitions and blogs. As a non-profit association, it also promotes photographic heritage through events and training. Its services include expert advice on digitization equipment, IPR issues, open access policies, and training on the MINT mapping tool.

EUscreen (Audiovisual Heritage) EUscreen is a consortium of European broadcasters and audiovisual archives that serves as Europeana's aggregator for television and AV content. It has contributed over 1 million records from 34 archives across 28 countries. In addition to its aggregation role, EUscreen runs its own portal (euscreen.eu). The consortium, established in 2006, offers services including curating video collections, promotion through its open-access VIEW Journal, and training on workflows and metadata mapping with the MINT tool. It also facilitates uploading, storage, and streaming of content, accepting data in XML, CSV, or JSON formats.

European Film Gateway (EFG) EFG is the aggregator for the film archive domain, representing 58 contributors from over 25 countries. It provides access to over 700,000 film historical documents, including photos, posters, and rare films. Like EUscreen, it also operates its own portal. EFG's services include guiding providers on metadata preparation and rights statements (which are particularly complex in the film sector), mapping native XML formats to the EFG and EDM schemas, and consolidating metadata by aligning local vocabularies with controlled EFG vocabularies.

MUSEU (Museum Collections) MUSEU is the accredited aggregator for European museums and institutions holding museum collections. Run by the MICHAEL Culture Association, it has provided Europeana with access to over 5.5 million records from 350 museums. Its comprehensive services include advice on digitization, data modeling, mapping, multilingual terminologies, and IPR. It provides both collective and tailored training, as well as access to storytelling tools. For data modeling, MUSEU typically implements the LIDO schema and uses the MINT mapping tool for transformation.

EUreka 3D (3D Content) EUreka 3D is a project funded by the Digital Europe Program to support the complex digital transformation required for 3D digitization. It addresses the need for cultural heritage institutions to modernize their processes, retrain personnel, and upgrade infrastructure to handle 3D content. The project's goal is to produce high-quality 3D digitization with metadata ready for harvesting to Europeana. To achieve this, it uses the EUreka3D Data Hub, which was expressly designed for Europeana compatibility. This hub includes a 3D visualization library embeddable in the Europeana website, an input form to describe metadata in EDM, a service to assign Persistent Identifiers (PIDs), and communicates with Europeana via OAI-PMH.

Archives Portal Europe This is the domain aggregator for archival material, maintained by the Archives Portal Europe Foundation. It provides access to millions of descriptive units, known as "finding aids," from thousands of archival institutions across Europe. It supports its providers by mapping various national archival standards (like EAD, EAC-CPF, and EAG) to the Europeana Data Model, offering data validation, conversion, and storage services.

Europeana Sounds As the domain aggregator for audio and audio-related content, Europeana Sounds is hosted by the British Library. It brings together music, spoken word, and soundscapes from archives and libraries. It offers partners extensive support, including advice on data quality, metadata mapping using the MINT tool (which it provides free access to), and guidance on copyright and licensing for audio-visual materials.

European Fashion Heritage Association (EFHA) EFHA is the thematic aggregator for fashion heritage. It works with over 45 public and private institutions, including museums, brand archives, and private collectors, to share high-quality digital content. Its services include digital curation (blogs, exhibitions), metadata enrichment, and training workshops. It is flexible in its technical requirements, accepting metadata via OAI-PMH, XML, or CSV.

Jewish Heritage Network (JHN) The JHN operates the Judaica Europeana aggregator, focusing on the rich tapestry of Jewish cultural heritage. It works with museums, libraries, and research centers to provide access to content. JHN offers its partners comprehensive support, including advice on metadata formats and mapping, content preparation, data storage, and specific metadata enrichment related to Jewish heritage.

Manuscriptorium Run by the National Library of the Czech Republic, Manuscriptorium is the domain aggregator for historical written resources, such as manuscripts, incunabula, and early printed books. It provides a specialized digital environment for these materials. Its services include advising partners on OAI-PMH and IIIF standards, data conversion, and data storage for contributors. It requires metadata to be provided in XML.

MIMO - Musical Instrument Museums Online MIMO is the domain aggregator for musical instruments, bringing together collections from public museums worldwide. It offers specialized services including data mapping and conversion to the LIDO (Lightweight Information Describing Objects) schema, data enrichment using its own multilingual thesaurus of instruments, and technical advice on digitization.

OpenUp! (focused on natural history) OpenUp! is the accredited aggregator for the natural history domain, connecting museums, botanical gardens, and research collections. It specializes in handling the specific metadata standards of the biodiversity community (like Darwin Core and ABCD). It provides a helpdesk, training, and a technical infrastructure based on GBIF and BioCASe tools to transform and map this specialized data for Europeana.

TIB - Leibniz Information Centre for Science and Technology and University Library (specifically its AV Portal) TIB serves as the thematic aggregator for scientific and technical audiovisual content through its AV Portal. It focuses on aggregating scientific videos, such as conference recordings, experiments, and instructional videos. TIB was accredited in 2020 and has been active in projects to enhance Europeana's infrastructure for AV media, ensuring that this specialized video content is discoverable.

8.2 Aggregation to Europeana: the complexities of the process

8.2.1 Challenges in Aggregation for 3D Content

Cultural heritage institutions aiming to publish online collections of 3D models and share them within the Europeana ecosystem face a range of technical and organizational challenges. These include storing files in multiple formats for different user groups, enabling online visualization for discovery, aligning metadata with the evolving Europeana Data Model, and providing detailed paradata about the digitization and modeling processes. Currently, many institutions address these needs in a fragmented way, relying on a mix of in-house and outsourced services. This often leads to duplicated efforts, inconsistent workflows, and complex orchestration, highlighting the need for more streamlined and coordinated solutions. The EUreka 3D project (in its D2.3) has been tasked to tackle several issues complicating the aggregation of 3D content to Europeana, such as:

- **Storing files and providing access to the model in different formats to relevant categories of users.** The formats used for the storage of 3D data and raw data may not be the best choice for visualisation or delivery of the model online, due to the size of these files (often quite large and not suitable to be transferred online) and to the processing of 3D data that is done on the client side (and thus the actual device used by the user plays a key role in the user experience of the 3D model). 2D content can also "suffer" from the same problems but they are more prominent in 3D because of its greater space requirements and overall complex structure. . To partially cope with the challenges, different versions of the same model can be made available by the CHIs to the different categories of users they deem more appropriate, based on the CHI's digitisation use case reusability assessment, via controlled or open access on a cloud.
- **Enabling visualisation of the model over the internet for online access and discovery in Europeana.** Displaying 3D models online depends heavily on the user's device, which affects performance due to factors like network speed and processing power. Despite advancements in the 3D industry, there's still no unified, open-access solution that connects 3D data processing with visualization tools. To view 3D models online, users need a viewer that allows interaction, but most available options are commercial, often non-European (e.g., Sketchfab), raising concerns about data security and long-term access. External viewers can cause issues like duplicated storage, workflow inefficiencies, and version mismatches. Moreover, few viewers are compatible with Europeana, making it difficult to embed 3D models directly into its records.
- **In addition to providing regular metadata information, reconciling metadata to the current and evolving structure of the EDM.** Aggregators like Photoconsortium help CHIs to comply with all EDM requirements. However,

the absence of an EU-based, integrated solution to all these challenges has created an evident need in the cultural heritage sector.

- **Providing access to the full set of paradata** containing information about the 3D digitisation and modelling process.
- **Providing access to different formats for different reuse cases** The formats used for the storage of 3D data and raw data may not be the best choice for visualisation or delivery of the model online, due to the size of these files (often quite large and not suitable to be transferred online) and to the processing of 3D data that is done on the client side (and thus the actual device used by the user plays a key role in the user experience of the 3D model). These are not intrinsic problems for 3D, as 2D content also suffers from them, but they are more prominent in 3D because 3D is more complex in nature, and 3D content requires extensively larger amounts of space than 2D content, which affects its storage, processing and transfer over a network. To partially cope with the challenges, different versions of the same model can be made available by the CHIs to the different categories of users they deem more appropriate, based on the CHI's digitisation use case reusability assessment, via controlled or open access on a cloud.

8.2.2 The Eureka3D datahub

The data management system and workflow developed specifically in EUreka3D is a comprehensive solution for CHIs to manage 3D assets and share them online. All the phases, from storage to visualisation to the addition of metadata and paradata are managed in the EUreka3D Data Hub, which also communicates with the Europeana Metis ingestion tool for harvesting via OAI-PMH. The intermediation of Photoconsortium as accredited aggregator to Europeana supports quality checks of the EDM-based metadata to comply with EPF. The development of the EUreka3D Data Hub has followed an iterative process that saw participation of all project partners, including content providers (who represent the category of users of the EUreka3D infrastructure), and Photoconsortium and Europeana (who supported the development of the aggregation service in the EUreka3D Data Hub to be fully integrated and compliant with Europeana technical and procedural frameworks). Content providers used the facilities of the EUreka3D Data Hub for storing the data in the cloud, both the raw data from the digitisation and the more refined models in different formats if applicable. The content providers had to register to the Check-in service and join the EUreka3D Community. This flexible system for granting different levels of access and editing rights to different communities is explained in detail in the D3.2 The EUreka3D AAI architecture. Storage and data management facilities in the EUreka3D Data Hub are supported by servers, virtual machines and compute power based in the EU, hosted by the national providers of the European Grid Initiative (EGI) Federation and specifically by Affiliate Entity Cyfronet. The user can upload the files individually or via an API service for batch uploading. The EUreka3D Data Hub also integrates an open-source 3D Viewer. The current version is a basic tool that enables visualisation of various formats of 3D models provided in a zip file. The viewer is fully compatible and embedded in Europeana. All the files uploaded in the EUreka3D Data Hub can be shared on the internet via a shared, open data tool which also enables users to request a PID, provided by the service of EUDAT B2HANDLE, in partnership with the EUreka3D project. The metadata for each object can be added either via XML or via a metadata input form, developed in close collaboration with Photoconsortium and Europeana to include all the mandatory and recommended fields to create a valid and rich EDM file. The metadata input form enables the user to add literal values and LOD links. Various elements and fields are automatically added in the metadata such as the PID, the link to the viewer (isshownby), the file size and others. In the metadata input form, it is possible to include a URL link that leads to the paradata report associated with the 3D model and a link to the raw data from the digitisation, available for downloading from the EUreka3D Data Hub. The dataset in EDM format is then shared with Europeana through the harvesting of information via OAIPMH, and after quality checks performed both manually and in the Metis Sandbox, eventually the record is published on Europeana.eu.

9 Recommendations on Metadata Simplification for Aggregation in MUVEs

Cultural heritage institutions of all sizes play a crucial role in shaping sustainable and interoperable digital ecosystems. From our practice-based, institutional perspective, metadata simplification is not merely a technical task, but a vital strategic approach that supports long-term preservation, accessibility, and reuse. As cultural heritage institutions increasingly engage with MUVEs and contribute to platforms like Europeana, several key recommendations emerge.

Notably, research conducted within the IMPULSE project, particularly the findings of D17 (Ongoing verification of the use of digital heritage objects within the emerging platforms) of Work Package 3 (WP3), has revealed a significant gap in the current digital infrastructure: there are virtually no established metadata standards or simplification guidelines specifically designed for MUVEs. This absence is striking, especially given the growing importance of these environments in cultural engagement, education, and digital heritage dissemination, especially in the context of emerging European Collaborative Cloud for Cultural Heritage. The lack of structured, interoperable metadata frameworks tailored to MUVEs not only hampers consistent documentation and discoverability but also poses long-term risks for data preservation and cross-platform integration.

Similarly, the findings of D11 (Report on the review of the latest MUVE technologies, processes, formats, best practices), impediments of WP2 address this significant gap in the current digital infrastructure. This absence is particularly problematic given the growing complexity and collaborative nature of these environments, which demand interoperable and user-friendly data practices. Moreover, the fragmentation of repositories and the heterogeneity of metadata schemas across platforms like Europeana further complicate content integration and reuse. The lack of standardised metadata structures and APIs limits discoverability, accessibility, and the potential for cross-platform collaboration. These challenges are compounded by the absence of simplification strategies that could support cultural heritage institutions in managing and contributing content to MUVEs. The findings underscore the urgent need for coordinated efforts to develop metadata standards and simplification recommendations that are specifically designed for MUVEs. Such efforts should be grounded in practice-based, iterative methodologies that reflect the real-world needs and constraints of cultural heritage institutions.

Additionally, D12 (Internal Report on the suitability of different types of content and their integrity based on analyzed cases with recommendations for the conversion of digitized files and post-processing steps) of WP2 confirms the existence of the persistent "cultural heritage paradox": despite extensive digitisation efforts, much of the resulting content remains underutilised due to systemic barriers in metadata, technical compatibility, and implementation readiness. A key finding was the "metadata crisis", a widespread lack of structured, standardised, and AI-ready metadata. This absence severely limits the discoverability, contextualisation, and reuse of digital assets. Even technically sophisticated digitised materials often lack basic descriptive, structural, or administrative metadata, making it difficult to integrate them into XR, gaming, or educational applications.

These findings underscore the urgent need for targeted standardization efforts and policy development that address the unique characteristics of MUVEs. Without such foundational work, cultural heritage institutions may struggle to ensure that their digital contributions remain accessible, reusable, and meaningful in increasingly immersive and collaborative digital spaces. As cultural heritage institutions will be increasingly engaging with MUVEs and contribute to platforms like Europeana or the future ECCCH, several key recommendations are proposed to mitigate the structural absence of metadata standards and simplification strategies in MUVEs. Our recommendations have emerged directly from a practice-based perspective that departs from the IMPULSE Consortium as a case study, and they have been shaped through iterative processes of experimentation, reflection, and institutional collaboration throughout the development of the platform, thus ensuring they are both grounded in real-world challenges and responsive to evolving digital contexts.

- **Persistent link to the original metadata**: Building upon the "metadata first" framework proposed by WP2, institutions managing rich metadata sets must, above all, ensure that a persistent link to the original metadata is always included. This practice supports provenance, enables traceability, and allows users and systems to verify and contextualize data across platforms and over time. Metadata should be designed with future preservation and interoperability in mind, aligning with open standards from the outset to remain usable in evolving technological contexts. Daily practices in cultural institutions show that a simple hyperlink to a human-readable webpage is fragile and of limited use for automated systems. For this strategy to be robust and future-proof, the link must be a Persistent Identifier (PID), such as a DOI or ARK, that resolves to a landing page offering machine-readable metadata endpoints (e.g., an API delivering RDF/XML or JSON-LD).
- **Use of lightweight standards**: We strongly advocate for the adoption of lightweight, open, and well-documented standards. These reduce implementation barriers, foster collaboration, and simplify workflows. Among these, Dublin Core stands out as a preferred option due to its simplicity, flexibility, and widespread adoption. It supports meaningful metadata exchange and embedding, even for institutions with limited technical resources, and aligns well with international standards.
- **Minimal mandatory set for Dublin Core:** As part of our ongoing efforts to address metadata fragmentation and enhance interoperability within the IMPULSE project, a collaborative decision was made, following a thorough review of the collection factsheets presented in D18 (Ground truth dataset for further usage within the project). We propose adopting a minimal set of mandatory metadata elements based on the Dublin Core standard. This shared schema will serve as a foundational layer for consistent documentation across partner institutions. The selected core fields are those of the **contributor** (the entity responsible for making contributions

to the resource), **description** (an account for the resource), **identifier** (an unambiguous reference to the resource within a given context)**, rights** (information about rights held in and over the resource)**, and title** (a name given to the resource**).** They were chosen for their broad applicability, semantic clarity, necessity, frequency of use in the cultural heritage ecosystem, and relevance to both preservation and implementation contexts, without excluding richer metadata sets. By establishing this common baseline, we aim to facilitate more effective integration of digitised cultural heritage assets into immersive environments and ensure their long-term discoverability and reuse.
- **Metadata being accessible from within the (IMPULSE) platform.** While this feature is not yet active on the IMPULSE platform, the system architecture allows metadata to be easily displayed and integrated, and its implementation is foreseen in upcoming development phases.
- **Integration of AI-enhanced methodologies.** Institutions should implement integrated workflows that can combine traditional expertise with AI-enhanced implementation procedures in MUVEs. Within IMPULSE, WP2 is conducting research on the integration of traditional and AI-enhanced workflows (D 2.2, Tech Internal Report 1).
- **Addressing the challenges**: However, metadata aggregation is not without challenges. Many institutions do not openly share the difficulties they face, which limits collective learning and the development of shared solutions. Our position emphasizes the importance of transparency and collaboration. By committing to a single lightweight standard, institutions can avoid fragmentation, reduce redundancies, and streamline training and implementation efforts.
- **Challenges and opportunities of Metadata in Relation to Emerging Technologies**: Emerging technologies are transforming metadata from a static descriptive layer into a dynamic enabler for simulation, AI training, and cross-domain interoperability. The opportunity lies in leveraging rich, multimodal metadata covering not only visual but also non-visible formats like multispectral, hyperspectral, and acoustic data, in order to support advanced modeling of physical properties and behaviors. This creates new value streams for institutions that curate cultural or scientific data, enabling multipurpose reuse across platforms. However, challenges include ensuring standardization across diverse ecosystems, maintaining semantic consistency, and addressing scalability and privacy concerns as metadata becomes more granular and computationally critical.

Currently, many cultural heritage institutions address metadata-related tasks and management in a decentralized manner, relying at times on a mix of in-house and outsourced services. This leads to duplicated efforts and complex workflow orchestration. To foster reuse and enrich the emerging Common European Data Space for Cultural Heritage, institutions must tackle key issues: providing files in multiple formats for diverse users, enabling online visualization for discovery, reconciling metadata with the evolving Europeana Data Model, and providing access to detailed paradata about the digitization and modeling process. By adopting a unified, forwardlooking approach to metadata simplification, heritage institutions can enhance the quality, accessibility, and sustainability of their digital collections, and ultimately contribute to a richer and more inclusive European cultural heritage landscape.

10 Concluding remarks

This deliverable originates from critical observation: the near-total absence of established metadata standards for virtual and immersive environments. As highlighted in previous IMPULSE Deliverables, this gap presents a significant barrier to the discoverability, interoperability, and long-term preservation of digital cultural heritage within Multi-User Virtual Environments (MUVEs). Despite the increasing relevance of immersive platforms in cultural engagement and cultural heritage institutions, the lack of structured, interoperable metadata frameworks tailored to these contexts has resulted in fragmented practices and under-utilized digital assets. In response, the IMPULSE Consortium has adopted a simplification strategy grounded in both analytical evaluation and institutional practice. At its core is the adoption of Dublin Core as a lightweight, widely accepted metadata standard, supported by a minimal set of mandatory fields. This approach enables institutions of all sizes and capacities to participate in the digital transformation of cultural heritage without the burden of complex or resource-intensive metadata models.

In this light, the IMPULSE platform is being developed with the capability to make metadata accessible and easily displayable. While this feature is not yet active, the system architecture has been designed to support seamless integration and visualization of metadata in future development phases. This ensures that the platform remains adaptable and aligned with the FAIR principles, while also anticipating the evolving needs of cultural heritage institutions and users.

Simplification is pivotal for IMPULSE because it lowers the entry barrier for cultural institutions, ensuring that metadata practices in MUVEs become more consistent and interoperable. By adopting pragmatic and widely recognized standards, the project sets the foundation for scalable integration and future enhancements that align with FAIR principles. Moving forward, this approach will guide the next development phases, enabling richer metadata visualization and facilitating a truly connected ecosystems European digital cultural heritage will be moving to the direction of Multi-user virtual environments.

Ultimately, this deliverable calls for a shift from fragmented, resource-intensive practices to a harmonized framework that will empower cultural institutions to participate meaningfully in the digital transformation of European heritage and contributes to the development of the Common European Data Space for Cultural Heritage.

11 Attachment 1: An overview of the metadata schemas and export file types used within each IMPLUSE institution and collection.

Jagiellonian University Collections

1.1 Collections of Art and Scientific Objects

Metadata Schema: Dublin Core

Export File Types: .txt, .xlsx, .csv

1.2 Alexander von Humboldt Collection

Metadata Schema: MARC21

Export File Type: RDF

1.3 Patrimonium Collection

Metadata Schema: MARC21

Export File Type: RDF

1.4 SLUBDRESEDN Collection

Metadata Schema: MARC21

Export File Type: RDF

Heritage Malta

2.1 Dockyard Collection

Metadata Schema: Dublin Core (extended with custom fields)

Export File Type: .csv

2.2 Maritime Collection

Metadata Schema: Dublin Core (extended with custom fields)

Export File Type: .csv

KU Leuven

3.1 Collectio Academica Antiqua

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.2 Corble Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.3 Glass Slide Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.4 Incunabula Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.5 Magister Dixit Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.6 Manuscripts and Manuscript Fragments Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.7 Picture Postcards Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.8 Jesuitica Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

3.9 Theses Collection

Metadata Schema: MARC21

Export File Types: .bib, .txt, .xlsx, .csv

Magna Żmien

4.1 Archives Collection

Metadata Schema: Essential Dublin Core, expanded with additional fields from CIDOC-CRM

Export File Types: .csv, .xlsx

4.2 Neolithic Cultural Heritage Collection

Metadata Schema: Essential Dublin Core, expanded with additional fields from CIDOC-CRM

Export File Types: .csv, .xlsx

National and Kapodistrian University of Athens

5.1 Scientific Instruments Collection

Metadata Schema: Descriptive metadata (schema name not mentioned)

Export File Type: .xlsx

5.2 History of NKUA – Interviews of the Athens University History Museum Building Inhabitants Collection

Metadata Schema: Descriptive metadata (schema name not mentioned)

Export File Type: .doc

5.3 Medical Sciences Collection

Metadata Schema: Descriptive and contextual metadata (schema uncertain)

Export File Type: .xlsx

5.4 Portraits Collection

Metadata Schema: Descriptive metadata (schema name not mentioned)

Export File Type: No information available

5.5 Registry Books Collection

Metadata Schema: Descriptive metadata (schema name not mentioned)

Export File Type: No information available

Thessaloniki Film Festival Digital Archive

6.1 Movie Star Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.2 Books Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.3 Digital Prints Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.4 Catalogues Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.5 First Shot Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.6 Hellafi Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.7 Magazines Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.8 Photos Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.9 Posters Collection

Metadata Schema: Dublin Core

Export File Type: SQL

6.10 Publications Collection

Metadata Schema: Dublin Core

Export File Type: SQL

Film University Babelsberg KONRAD WOLF

7.1 Volumetric Contemporary Testimony of Holocaust Survivors Collection

Metadata Schema: Dublin Core (tentative)

Export file type: tentative

Film Museum Potsdam

8.1 Costume Design and Scenography Collection

Metadata Schema: Dublin Core (tentative)

Export file type: tentative

8.2 Film- Cinema Technology Collection

Metadata Schema: Dublin Core (tentative)

Export File Type: tentative

8.3 Props Collection

Metadata Schema: Dublin Core (tentative)

Export File Type: tentative

Film University Babelsberg and the Film Museum Potsdam are currently actively reorganizing their collections, recording and annotating the testimonies. The Museum currently operates with their own internal archival format and is in the process of restructuring the collection data. The envisioned schema for all the collections is going to be Dublin Core.

12 Attachment 2: Proposed Set of Mandatory Dublin Core Elements.

Term name	Definition	Comment	Mandatory
Contributor	An entity responsible for making contributions to the resource.	Examples of a Contributor include a person, an organization, or a service. Typically, the name of a Contributor should be used to indicate the entity.	x
Coverage	The spatial or temporal topic of the resource, the spatial applicability of the resource, or the jurisdiction under which the resource is relevant.	Spatial topic and spatial applicability may be a named place or a location specified by its geographic coordinates. Temporal topic may be a named period, date, or date range. A jurisdiction may be a named administrative entity or a geographic place to which the resource applies. Recommended best practice is to use a controlled vocabulary such as the Thesaurus of Geographic Names [TGN]. Where appropriate, named places or time periods can be used in preference to numeric identifiers such as sets of coordinates or date ranges.	
Creator	An entity primarily responsible for making the resource.	Examples of a Creator include a person, an organization, or a service. Typically, the name of a Creator should be used to indicate the entity.	
Date	A point or period of time associated with an event in the lifecycle of the resource.	Date may be used to express temporal information at any level of granularity. Recommended best practice	

		is to use an encoding scheme, such as the W3CDTF profile of ISO 8601 [W3CDTF].	
Description	An account of the resource.	Description may include but is not limited to: an abstract, a table of contents, a graphical representation, or a free-text account of the resource.	x
Format	The file format, physical medium, or dimensions of the resource.	Examples of dimensions include size and duration. Recommended best practice is to use a controlled vocabulary such as the list of Internet Media Types [MIME].	
Identifier	An unambiguous reference to the resource within a given context.	link to the object on the contributor's platform	x
language	A language of the resource.	Recommended best practice is to use a controlled vocabulary such as RFC 4646 [RFC4646].	
Publisher	An entity responsible for making the resource available.	Examples of a Publisher include a person, an organization, or a service. Typically, the name of a Publisher should be used to indicate the entity.	
Relation	A related resource.	Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.	

Rights	Information about rights held in and over the resource.	Legal and terms and conditions	x
Source	A related resource from which the described resource is derived.	The described resource may be derived from the related resource in whole or in part. Recommended best practice is to identify the related resource by means of a string conforming to a formal identification system.	
Subject	The topic of the resource.	Typically, the subject will be represented using keywords, key phrases, or classification codes. Recommended best practice is to use a controlled vocabulary.	
Title	A name given to the resource.	info about the object	x
Type	The nature or genre of the resource.	Recommended best practice is to use a controlled vocabulary such as the DCMI Type Vocabulary [DCMITYPE]. To describe the file format, physical medium, or dimensions of the resource, use the Format element.	

IMPULSE

IMmersive digitisation: uPcycling cULtural heritage towards new reviving StratEgies

Funded by the European Union