GA – PROJECT NUMBER:	**101158046**
PROJECT ACRONYM:	AUTOMATA
PROJECT TITLE:	AUTOmated enriched digitisation of Archaeological liThics and cerAmics
CALL/TOPIC:	HORIZON-CL2-2023-HERITAGE-ECCCH-01-02
TYPE OF ACTION	HORIZON RIA
PRICIPAL INVESTIGATOR	Prof Gabriele Gattiglia, UNIPI
TEL:	+39 050 2215228
E-MAIL:	gabriele.gattiglia@unipi.it

This project has received funding from the European Union's HORIZON RIA research and innovation programme under grant agreement N. 101158046

D 5.1 Ontology and Metadata scheme for enriched digitisation

Version: 1.0

Revision: first release

Work Package: 5 - Technologies for enriched digitisation
Lead Author (Org): Xavier Granier (Institut d'Optique Graduate School / LP2N)
Contributing Author(s) (Org): Julian D. Richards (UoY), Arthur Leck (UBM), Rémy Chapoulie (UBM), Sarah Tournon (UBM), Bruno Dutailly (UBM), Martina Naso (UNIPI), Gabriele Gattiglia (UNIPI).
Due Date: M10
Date: 30/06/2025

Project co-funded by the European Commission within the ICT Policy Support Programme		
Dissemination Level		
P	Public	X
C	Confidential, only for members of the consortium and the Commission Services	

Revision History

Revision	Date	Author	Description
0.1	19/06/2025	Xavier Granier	First document
0.2	23/06/2025	Xavier Granier, Arthur Leck, Rémy Chapoulie, Sarah Tournon, Bruno Dutailly	Section 1: draft content added. Section 2: start.
0.3	24/06/2025	Xavier Granier, Arthur Leck, Rémy Chapoulie, Sarah Tournon	Section 2: draft content added. Section 3: start.
0.4	26/06/2025	Xavier Granier, Julian D. Richards, Arthur Leck, Sarah Tournon	Section 4: reorganisation, corrections
0.5	27/06/2025	Xavier Granier, Julian D. Richards, Arthur Leck, Sarah Tournon	Summary. Improved consistency with AO-Cat. Minor changes in naming (Figure redone)
0.6	29/06/2025	Martina Naso, Xavier Granier	Revision of the document. Last naming change for Digital Container
0.7	30/06/2025	Gabriele Gattiglia, Martina Naso	Content added and final revision.

Disclaimer

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Abbreviations

WP: Work package

M: Month

UNIPI: Università di Pisa

UBM: Université Bordeaux Montaigne

UoY: University of York

INRAP Institut National de Recherches Archéologiques

AMZ: Arheoloski Muzej u Zagrebu

QB: QBrobotics Srl

HUJ: The Hebrew University of Jerusalem

MIN: Miningful srls

KCL: King's College London

IIT: Fondazione Istituto Italiano di Tecnologia

UB: Universitat de Barcelona

CL: Culture Lab

Executive summary

This deliverable outlines the metadata scheme designed for the AUTOMATA project, which aims to deliver enriched digitisation of archaeological artefacts through automated systems. In line with the FAIR principles (Findable, Accessible, Interoperable, Reusable) and the broader goals of Open Science, this metadata scheme ensures that all data produced are traceable, interoperable, and preserved in the long term.

Crucially, the metadata structure serves a core purpose: enabling the creation and reproducibility of 3D models for each artefact, while spatially linking these models to archaeometric data acquired during the process. To achieve this, every acquisition must be described in detail, every dataset must be properly stored, and each must be easily linked to the spatial positions of both the artefacts and the sensors that captured them.

The proposed scheme builds on established standards such as CIDOC-CRM and incorporates best practices from initiatives like ARIADNEplus. It captures essential metadata — including actors, technologies, dates, locations, physical objects, deposits, and processes — while adapting to the specificities of automated acquisition workflows. The resulting structure ensures not only that the necessary information is present, but that meaningful links can be maintained between all elements, especially between sensor data and spatial context. It also promotes the use of persistent and unique identifiers and controlled vocabularies to support integration with European research infrastructures and facilitate long-term data reuse.

1 Context and state of the art

The AUTOMATA project was funded under the European Collaborative Cloud for Cultural Heritage call, which is being constructed under the auspices of the leading project ECHOES[1]. One of its primary goals is to make scientific and cultural data available in a transparent way in the spirit of Open Science. The metadata structure outlined in this deliverable explicitly aligns with ECCCH's objectives, ensuring compatibility with broader European digital heritage initiatives and infrastructures.

1.1 FAIR principles and metadata

The key principles to follow in order to respect the goal of Open Science have been formalised under the acronym **FAIR**: the data must be **F**indable, **A**ccessible, **I**nteroperable, and **R**eusable (Wilkinson et al., 2016). To achieve such goals, metadata schemes and ontologies are playing an essential part. Their power of description makes them more easily **F**indable than raw data alone by associating them with keywords, dates, authors, and so on. By harmonising metadata schemes through mapping ontologies and vocabularies to open data standards, they improve **I**nteroperability. They participate in the **R**eusability by providing all the complementary information and context that helps in interpreting them.

As pointed out in Richards (2023), "there is certainly a risk that a network of repositories operating at national level creates a series of data silos. In fact, digital repositories need to do two things: they must bring resources together and make it easy for users to interrogate them via shared and user-friendly interfaces, but they must also open data up: via APIs, harvesting protocols such as OAI-PMH, and as Linked Open Data (LOD) so that it can be viewed and manipulated by multiple routes (May, Binding, & Tudhope, 2015)". This is the core idea of **I**nteroperability.

Interoperability arises from one part with respect to common or compatible standards for exchanging that information. The search for a perfect representation of such metadata is far from new. It must balance the need for the best descriptive information with the compatibility in a multidisciplinary world. The now-standard and minimal approach is the Dublin Core (DCMI Usage Board, n.d.) Metadata Element Set. However, as a lowest common denominator, it may lose a lot of specific but pertinent information, particularly since it is a flat file model that does not preserve the relationships between entities.

1.2 For cultural heritage and 3D

The ARIADNEplus project (Salas Rossenbach, 2022) was a keystone in the development of a solution dedicated to Archaeology and Cultural Heritage. ARIADNE defined an ontology - the AO-Cat (Felicetti, Meghini, Richards, & Theodoridou, 2025) - which underpins its online portal and facilitates searching by the powerful combination of "what", "when, and "where". The AO-Cat is a subset of the CIDOC-CRM ISO standard ontology (Doerr, Ore, & Stead, 2007). It is an event-based data model that also provides a way of describing and relating digital archaeological resources of

[1] ECHOES (European Cloud for Heritage OpEn Science): https://www.echoes-eccch.eu/

all data types, including textual, images, GIS, 3D, and others. The AO-Cat was selected and further enriched to meet the needs of the archaeological and cultural heritage domains, including the description and integration of 3D data. ARIADNE partners, such as ADS[2], gain interoperability and findability by mapping their native data schemas to the AO-Cat.

As part of the European Research Infrastructure Consortium (ERIC) DARIAH-EU[3], the MASA and 3D consortia of the research infrastructure Huma-Num[4] have also proposed tools and ontologies. Inside the ARIADNEplus project, these projects have been able to map their own data structures to the AO-Cat. Indeed, the transposition of good practices, even if differently organised, leads to the same core information, thus facilitating the aggregation of archaeological datasets at an international scale (Richards, 2023). As an example, the Consortium 3D for Digital Humanities (C3DHN)[5] has developed its own metadata scheme dedicated to 3D data for Human and Social Sciences (Quantin et al., 2023) used in the 3D Data Repository for Humanities (3DRH)[6]. It is mapped to CIDOC-CRM to be harvested in ARIADNEplus through the platform OpenArcheo (Marlet et al., 2019).

Recently, a mapping to the Europeana Data Model (Clayphan, Charles, & Isaac, 2016) has been demonstrated (Quantin et al., 2025) in the context of the Eureka3D project (Opiola & Lamata Martinez, 2024).

[2] ADS (Archaeology Data Service): https://archaeologydataservice.ac.uk/
[3] DARIAH-EU (Digital Research Infrastructure for the Arts and Humanities): https://www.dariah.eu/
[4] Huma-Num Consortia: https://www.huma-num.fr/les-consortiums-hn/
[5] C3DHN (Consortium “3D for Digital Humanities”): https://3dhumanities.hypotheses.org/category/publications
[6] 3DRH (3D Data Repository for Humanities): https://3d.humanities.science/

2 Proposal: general metadata

2.1 Introduction

Rich metadata schemes or ontologies based on established best practices have been shown in the previous Section to result in easily interoperable solutions. Consequently, the aim of the AUTOMATA project is not to introduce a new metadata scheme or ontology, but rather to encourage alignment with existing standards. The primary recommendation is to adopt the metadata schemes already in use within the two targeted infrastructures (ADS and 3DRH).

A graphical representation has been provided (Quantin et al., 2025) to facilitate understanding of the metadata output expected from the AUTOMATA project for the relevant data. The forthcoming section highlights the changes applied, primarily terminological, in order to adapt the structure for a broader audience beyond the 3D domain.

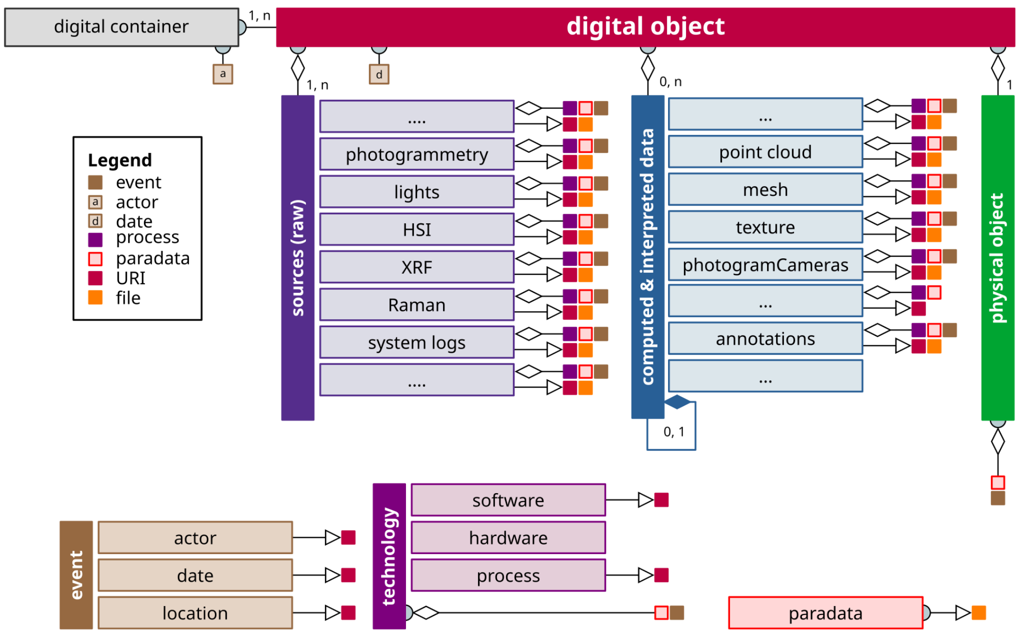

Fig. 1. Graphical overview of the required metadata for the AUTOMATA project, considering the use of photogrammetry for 3D creation, inspired by Quantin et al. (2025). Each block will be detailed in the document. 1,n stands for at least one element, while 0,n indicates that the number is free. 1 indicates that a unique element is mandatory.

Compared to the structure proposed in Quantin et al. (2025), and as illustrated in Fig. 1, few changes have been introduced to enhance compatibility and shared understanding with the AO-Cat ontology (Felicetti, Meghini, Richards, & Theodoridou, 2025):

- The class “deposit” will be renamed “digital container”, and is further detailed in Section 2.7.
- The class “3DObject” will be generalised as “digital object”, and is discussed in Section 3.1.

As highlighted by Felicetti et al. (2025), effective resource discovery requires support for the following types of search criteria:

- "What": searches based on topics, represented in the class "physical object" (see Fig. 4 and Section 2.5).
- "When": searches based on temporal criteria, addressed through the notion of "event" (see Fig. 2 and Section 2.3).
- "Where": searches based on spatial criteria, also incorporated into the notion of "event".

To further clarify the metadata requirements, two additional guiding questions have been introduced:

- "Who" is the creator, editor, or operator of the data and of the studied objects (modelled within the "event" class; see Fig. 2 and Section 2.3).
- "How" were the data and objects created (represented through the "technology" class; see Fig. 3 and Section 2.4).

2.2 Thesauri

To ensure interoperability, it is essential to rely on standardised descriptions. The adoption of the following controlled vocabularies and web services is recommended:

- The Getty Art and Architecture Thesaurus (AAT) has been adopted within the ARIADNE project to standardise the description of object and subject types. Partners already using their own vocabularies are not required to adopt the AAT directly, as these can be mapped to it, which serves as a common reference framework (Salas Rossenbach, 2022, Appendix 6). For instance, the French thesaurus PACTOLS (Nouvel et al., 2021) has been aligned with the AAT and can therefore be used within this shared structure.
- GeoNames[7] should be used to identify geospatial coordinates corresponding to modern place names.
- The Periodo[8] web service should be employed to define historical, art-historical, and archaeological periods. These are expressed in terms of absolute date ranges, linked to period names within specific geographic contexts and provided by recognised data providers.
- Each individual should be identified by an ORCID iD[9]. Although VIAF identifiers[10] are used in some cases (e.g. Quantin et al., 2025), ORCID is preferable, as an ORCID iD automatically generates a corresponding VIAF ID, whereas the inverse does not occur.

2.3 Importance of URI

As noted in Section 1.1, interoperability can be enhanced through the use of Linked Open Data. Linked Data establish typed connections between data from diverse sources by employing persistent identifiers (Bizer, Heath, & Berners-Lee, 2009). This is commonly achieved through Uniform Resource Identifiers (URIs), which uniquely identify resources on a specified server. A

[7] Geonames: https://www.geonames.org/
[8] Periodo: https://perio.do
[9] ORCID: https://orcid.org/
[10] VIAF: https://viaf.org/

well-known example is the Digital Object Identifier (DOI), which enables access to publications via links such as https://dx.doi.org/.

This approach allows data to remain distributed across multiple servers rather than being centralised, promoting accessibility in line with cloud-based principles. Accordingly, within the metadata scheme, data can either be embedded directly for known types or referenced externally as a local file or URI. It is therefore recommended to generalise the use of URIs.

To this end, each data and dataset that will be directly accessible on-line must be associated with a URI.

2.4 Event (Who/When/Where)

As shown in Fig. 2, an event, inspired by the CIDOC-CRM ontology, is characterised by three key aspects: WHO, WHEN, and WHERE. For example, regarding an acquisition event, this means identifying who was responsible for or oversaw the acquisition, as well as when and where it occurred. The use of the standardised thesauri described previously is recommended to describe these aspects consistently. The WHAT element, which inspired the AO-Cat ontology (Felicetti, Meghini, Richards, & Theodoridou, 2025), corresponds to the notion of "physical object" discussed in Section 2.5.

A note on terminology: although the three aspects WHO, WHEN, and WHERE are sometimes informally referred to as a "triplet" of information, this should not be confused with an RDF triple, which is a specific subject–predicate–object expression used in semantic web technologies. Here, the term simply denotes three distinct descriptive categories of an event.

In the AUTOMATA project, automation of acquisition processes is a key feature. This does not imply that the actor responsible for the acquisition (see Fig. 2) is the robot; the operator(s) controlling the robot remain the actual actors.

In some cases, such as with legal entities, an actor may be an institution. In these cases, a VIAF ID is more appropriate than an ORCID.

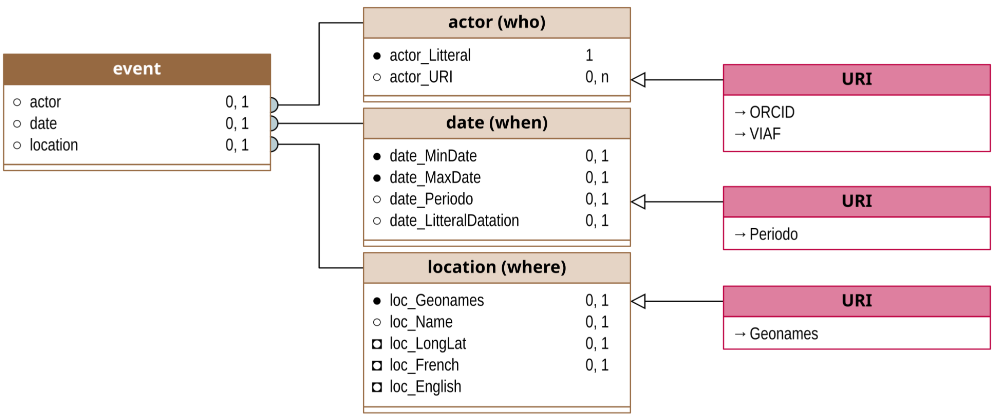

Figure 2. Graphical view of the notion of event. 0,1 indicates that there can be more than one value, 1 that a unique value is mandatory and 0,n that the number of values is free.

2.5 Technology (How)

For scientific reproducibility and interpretability of acquired data, it is mandatory to describe HOW the data have been acquired. Three key elements must be documented:

- Software: if the software source code is available, it is recommended to deposit it in Software Heritage[11]. Otherwise, all necessary information to identify the software and its version must be provided.
- Hardware: as a key component of AUTOMATA, each sensor must be referenced and all hardware characteristics described. These descriptions and the parameters used can be saved in a separate file as paradata. Calibration information must also be provided similarly. For further details on the sensors used in the project, refer to deliverable D2.1 (*Methodologies, Scenario, and User Requirements*).
- Process: the acquisition process must be described and it is recommended to publish these descriptions in the European Infrastructure SSH OpenMarket Place[12].

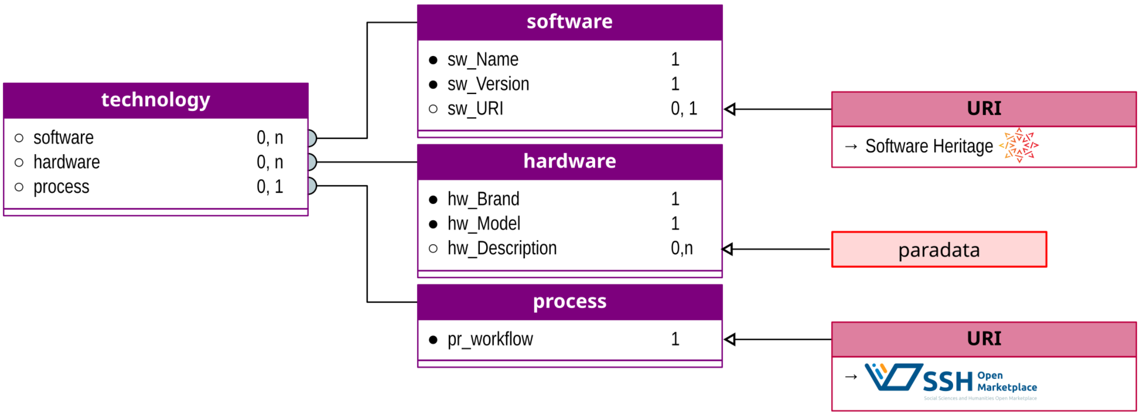

Figure 3. Graphical view of the notion of technology. 0,1 indicates that there can be more than one value, 1 that a unique value is mandatory; 0,n that the number of values is free.

2.6 Physical object (What or archaeological artefact)

This concept represents a core requirement of CIDOC-CRM ontologies applied to archaeology. It refers to the description of an artefact or site that either exists or has existed. All acquired and computed data constitute digital traces of a physical object. Within the AUTOMATA project, this refers specifically to the documentation of each sample to be digitised and analysed.

As illustrated in Fig. 4, each object, i.e., each sample digitised for the AUTOMATA project, must be identified by an inventory number and a name.

[11] Software Heritage: https://www.softwareheritage.org
[12] SSH OpenMarket Place: https://marketplace.sshopencloud.eu/

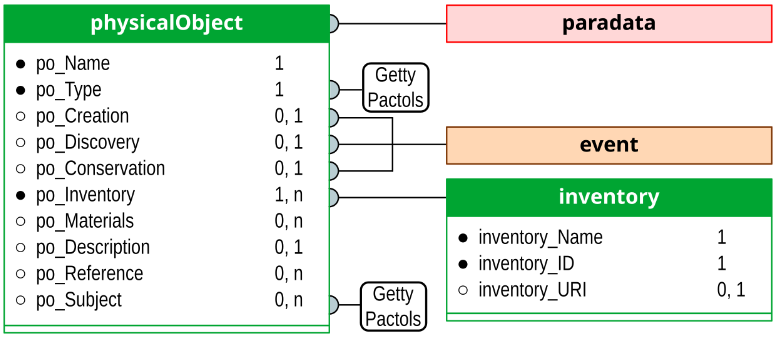

Figure 4. Graphical view of the notion of a physical object. 0,1 indicates that there can be more than one value, 1 that a unique value is mandatory; 1,n that at least one value is required and 0,n that the number of values is free.

2.7 Digital Container

A "digital container" refers to the structure used for publishing acquired data. Its content must be openly accessible and include at least one acquired dataset corresponding to a physical object. The container may also group together data related to multiple objects. Each container is associated with a DOI. In the context of the AUTOMATA project, a container may gather all automatically acquired data produced by the system for a given collection (see Fig. 1). The essential metadata elements are summarised in Fig. 5.

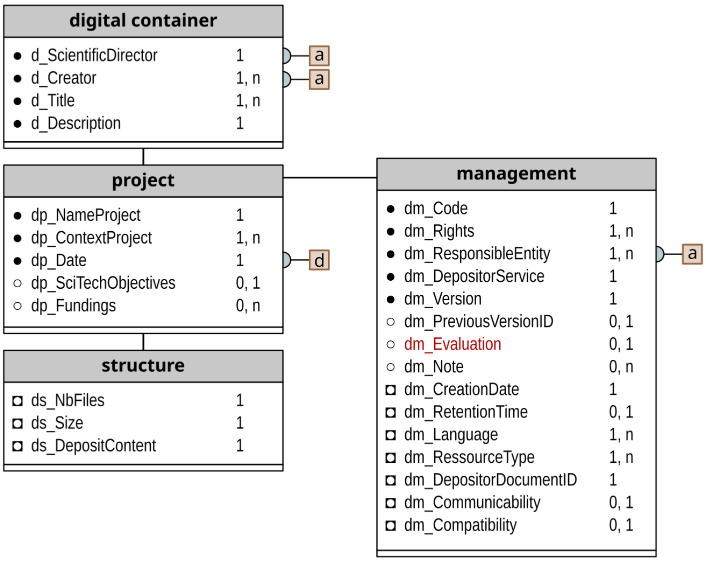

Figure 5. Graphical view of required information for a deposit 0,1 indicates that there can be more than one value, 1 that a unique value is mandatory; 1,n that at least one value is required and, 0,n that the number of values is free. a stands for actors and d for date (see Fig. 2).

3 For the Data

3.1 Digital Object

The digital object, building on the concept of the 3dObject described by Quantin et al. (2025), serves as the repository for all acquisitions. As illustrated in Fig. 1 and Fig. 5, a digital object represents the association between a physical object and its digital trace(s). A digital trace may consist of either source or raw data (see Section 2.3) or computed and interpreted data (see Section 2.4).

It is important to emphasise that, within the AUTOMATA project, a digital object is linked to a unique physical object but encompasses all acquired and computed data generated during an acquisition session. Fig. 5 provides an overview of all mandatory and optional associated metadata fields.

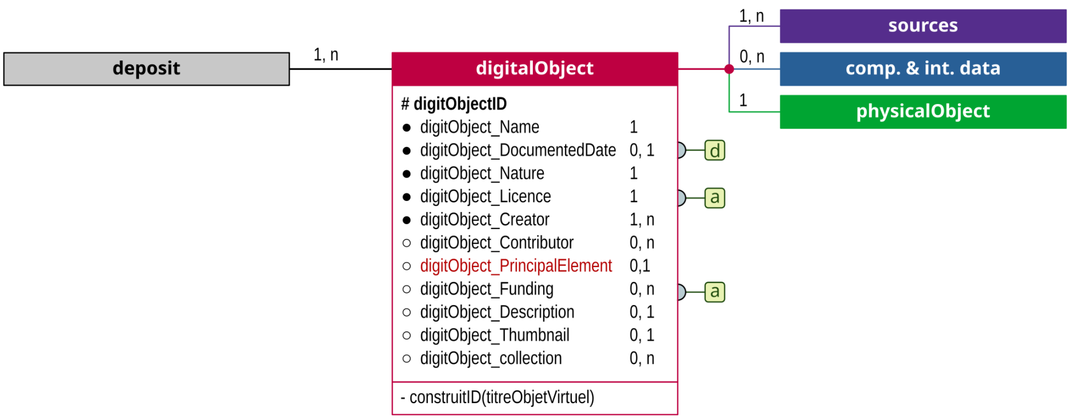

Figure 6. Graphical view of the notion of digital object. 0,1 indicates that there can be more than one value, 1 that a unique value is mandatory; 1,n that at least one value is required and, 0,n that the number of values is free. a stands for actors and d for date (see Fig. 2).

3.2 Sources or raw data

Sources comprise all existing resources, whether digital or physical, that predate the acquisition and have been used during the acquisition process. Raw data refer to the direct outputs from sensors, unmodified by either human or software intervention. These raw data must be provided alongside all relevant parameters and calibration information applied during their capture.

Raw data may originate from a range of 3D acquisition methods, such as photogrammetry, structured light scanning, or laser scanning. The proposed metadata schema is designed to be method-agnostic, provided that all essential metadata, including information on acquisition context, equipment, and configuration, is accurately recorded. The various types of raw data expected from the different devices used in AUTOMATA are detailed herein. This list is not exhaustive and may be supplemented by forthcoming deliverables describing the complete system.

Crucially, each acquisition within AUTOMATA must be spatially localised in 3D. To achieve this, all contextual data available at the time of measurement, including the positions of robots, sensors, and other hardware, must be captured and persistently linked to the raw outputs. Calibration data for robots and sensors, as well as the parameters and protocols employed, must also be documented. The system must ensure that each device provides the necessary metadata without omission of any critical information.

The system is required to output the following data and metadata throughout the acquisition process. It should be noted that the responsible agent (even in ontological terms) is the operator(s), not the system or robot itself.

- **System logs (robotic arm(s) and components):** a specific output of the AUTOMATA system consists of the robotic arms' parameters recorded at the time of each acquisition (e.g., translation, rotation, measurement triggering time). These logs must be recorded sequentially with each element clearly identifiable, with timestamps being mandatory. This information supports the computation of the 3D model and the precise localisation of each acquisition.
- **Lighting:** metadata related to lighting conditions, such as position, intensity, and status (on/off), must be recorded for each acquisition when changes occur. Each acquisition sequence should be associated with a unique ID.
- **3D model creation:**
  - **Photogrammetry:** raw outputs for photogrammetry, specifically sets of raw images, must be supplied without any modifications. If EXIF information is embedded, it must be preserved. Metadata should include the capture time to synchronise image acquisition with the robot's movement sequence. Paradata and metadata must also include all information necessary for manual or automatic rescaling of the resulting 3D model, referred to as topographic information. Each acquisition must be associated with a unique ID corresponding to the acquisition sequence.
  - **Structured Laser Scanning**: raw outputs from laser scanning must be provided unaltered, including detailed metadata on scanner settings, resolution, and acquisition parameters. Calibration data specific to the scanner must also be included to facilitate the accurate and reproducible generation of models. Each scan must have a unique identifier linked to the corresponding acquisition event.
- **Sensors (HSI/XRF/Raman):** metadata and/or paradata complementing the technology ontology must be provided by each sensor to aid in their positioning. Each acquisition must be linked to a unique ID corresponding to the acquisition sequence. This information, combined with robotic arm data, facilitates the most accurate possible estimation of measurement positions. Further details will be provided in Demonstrator 4.2 (*Final integration of sensing and acquisition systems*).

3.3 Computed or Interpreted data

This section addresses two primary types of data in the AUTOMATA project: 3D models and the 3D positioning of each measurement (including annotations or Referenced Information System in 3D -

RIS3D) (Dutailly, Portais, & Granier, 2023). For user-assisted generation, archiving and publishing these data is mandatory; however, in the case of automatic generation, this requirement is not compulsory.

Nonetheless, to facilitate the reusability of an acquisition sequence, access to the 3D positioning data is strongly recommended, as it prevents the need for recomputation, thus increasing its importance.

For 3D objects (whether point clouds or meshes), the following metadata must be considered:

Mandatory:

- Event: identification of who, when, and where the 3D object was created.
- Technology: description of how the 3D object was generated.
- File format: which can be extracted automatically from the file itself.
- File or URI reference.
- Unit: measurement units used for 3D positions (meters, centimetres, etc.). It is recommended to standardise all measurements to meters.
- Dimensions (X/Y/Z): scale along each dimension, representing the object's size.

Optional:

- Up axis: the axis designated as the vertical direction.
- Version: version number if multiple iterations have been produced.

If textures have been generated to provide colour for the 3D object, they are considered a distinct item. Indeed, for a single 3D object, multiple versions may exist due to variations in resolution, different software used for generation, or changes in parameters. When only the textures change, while texture coordinates remain the same, it is efficient to factorise the data. This referencing between computed elements is represented by the lower link (0,n) in Fig. 1.

During the 3D modelling process, whether based on photogrammetry or structured light scanning, all images or scans are aligned in 3D by computing a digital camera model for each image, or by referencing the scanner's spatial calibration data. Although these models may be estimated automatically, it is advisable to save them. These models are essential for registering all data in 3D, as proposed in the AIOLI approach (Abergel, Manuel, Pamart, Cao, & De Luca, 2023).

Regarding annotations, further details will be provided in forthcoming deliverables (D5.2, *3D Database*, and D5.3, *Reference Enriched 3D Data*). At the time of writing, it is recommended to adopt an approach compatible with the Web Annotation Data Model (WADM) (Sanderson, Ciccarese, & Young, 2017), which has been standardised for images by the IIIF consortium (Appleby et al., 2018). A working group is currently developing its 3D extension[13], which will be implemented in the Smithsonian Voyager project[14]. Additionally, within the consortium responsible for the

[13] IIIF 3D Community Group: https://iiif.io/community/groups/3d/
[14] Smithonian Voyager: https://smithsonian.github.io/dpo-voyager/

ECHOES project, the ATON Framework initiative (Fanini et al., 2021) is under development. It is noteworthy that all these solutions share common foundations with RIS3D, such as the use of JSON to describe 3D annotations.

4 Integration into the AUTOMATA Digitisation Sequence

To facilitate a clear understanding of how metadata generation is integrated within AUTOMATA, a detailed account of the information that must be recorded or entered for a complete digitisation sequence has been provided.

4.1 Before Starting the Digitisation Sequence

Before initiating the digitisation sequence, essential foundational metadata must be recorded. Primarily, this includes all common metadata identified in Section 2.7. Additionally, information that applies universally to all data is required.

Common Event
These metadata pertain to the entire acquisition session and apply to all source data:

- Who: the operator(s) (entered manually).
- When: start and end times of data acquisition (automatically recorded).
- Where: location of acquisition (entered manually).

Technology (System-Level Metadata)

- A general description of the system used.
- Sample preparation protocols.
- Configuration details (paradata), including calibration data, system parameters, etc.

Physical Object(s)
Each artefact must be described in a standardised manner:

- Inventory number and conservation location.
- Name.
- Type, using a standardised thesaurus (e.g., Getty / Pactols).
- Creation event, if known (date, location, and/or author).
- Discovery event, if known (date, location, and/or author).

Once these metadata elements are defined, the acquisition sequence can commence.

4.2 During the Acquisition Sequence

During the acquisition process, data and metadata are continuously recorded. Depending on their nature, these are categorised as either **source (raw) data** or **computed and interpreted data**.

System Logs (source/raw data)

- All changes to system parameters are logged.
- Each acquisition action is assigned a unique ID, which is recorded in the system's internal log.

This "system log" is crucial for linking information about robot movements and acquisitions at a later stage.

Acquisition Metadata for Each Measurement (source/raw data)

- Technology and method used (process and paradata).
- Resulting raw data (file).
- Timestamp of the event (automatically recorded).

On-the-fly Computation (computed and interpreted data)

- Description of the computation method (process and paradata).
- Resulting data (file).
- Timestamp of the event.

4.3 After the Acquisition Sequence

Once the acquisition sequence is complete, any further data processing, re-interpretation, or transformation must also be documented and stored within the computed and interpreted data section.

At the conclusion of the process:

- An archive containing all data and associated metadata should be created.
- This archive must be validated to ensure all required information is present.
- It can then be submitted to the chosen web service for long-term preservation.

Although much of the data and metadata recording will be automated, if metadata validation fails, adding the missing information can be a challenging task due to the volume of data involved. Therefore, a tool to assist with this completion is essential. For 3D objects, aLTAG3D (Dutailly et al., 2023) serves this purpose. It is an open-source software tool that provides a simple and efficient user interface for verifying, completing, and packaging metadata and data together for final submission.

5 Conclusions

This deliverable has defined the metadata structure necessary for the publication and long-term preservation of data produced within the AUTOMATA project. Particular emphasis has been placed on the balance between metadata that can be generated automatically and those requiring manual input, providing a foundation for integrating enriched digitisation into automated workflows.

The outlined approach serves as a practical guideline to support the refinement and development of the digitisation system. By adhering to existing standards and aligning with reference ontologies such as AO-Cat and CIDOC-CRM, the scheme promotes interoperability, reusability, and integration with European research infrastructures.

Rather than proposing a new model, the methodology presented reflects real-world needs and seeks to enhance the robustness and transparency of data production. The structure facilitates the accurate linking of 3D models with their corresponding physical artefacts and archaeometric data, enabling traceability and reproducibility at every stage of the acquisition process.

In doing so, this work contributes not only to the success of the AUTOMATA project but also to the broader objectives of the European Collaborative Cloud for Cultural Heritage. It provides a concrete foundation on which future recommendations, such as those developed through the ECHOES initiative, may be built. By adhering strictly to ECCCH's core principles of openness, interoperability, and sustainability, the proposed metadata scheme serves as a foundational building block for integrating archaeological digitisation workflows into a shared European digital infrastructure.

References

Abergel, V., Manuel, A., Pamart, A., Cao, I., & De Luca, L. (2023). Aïoli: A reality-based 3D annotation cloud platform for the collaborative documentation of cultural heritage artefacts. *Digital Applications in Archaeology and Cultural Heritage, 30*, e00285. https://doi.org/10.1016/j.daach.2023.e00285

Appleby, M., Crane, T., Sanderson, R., Stroop, J., & Warner, S. (2018). *IIIF Open/Web Annotation Extensions (V1.1)*. IIIF (International Image Interoperability Framework). Retrieved June 25, 2025, from https://iiif.io/api/annex/openannotation/

Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data: The story so far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22. https://doi.org/10.4018/jswis.2009081901

Clayphan, R., Charles, V., & Isaac, A. (2016, November). *Europeana Data Model – Mapping guidelines v2.4*. Europeana. https://pro.europeana.eu/page/edm-documentation

DCMI Usage Board. (n.d.). *Dublin Core™ Metadata Element Set, Version 1.1: Reference Description*. Retrieved June 23, 2025, from https://www.dublincore.org/specifications/dublin-core/dces/

Doerr, M., Ore, C.-E., & Stead, S. (2007). The CIDOC Conceptual Reference Model - A new standard for knowledge sharing. In *Proceedings* (pp. 51–56). https://doi.org/10.13140/2.1.1420.6400

Dutailly, B., Portais, J.-C., & Granier, X. (2023). RIS3D, a referenced information system in 3D. *Journal on Computing and Cultural Heritage*, 15(4), 73:1–20.

Dutailly, B., Tournon, S., Chayani, M., Grimaud, V., & Granier, X. (2023). *aLTAG3D: A user-friendly metadata documentation software*. Retrieved June 25, 2025, from https://hal.science/hal-04115897

Fanini, B., Ferdani, D., Demetrescu, E., Berto, S., & d'Annibale, E. (2021). ATON: An open-source framework for creating immersive, collaborative and liquid web-apps for cultural heritage. *Applied Sciences, 11*(22), 11062.

Felicetti, A., Meghini, C., Richards, J., & Theodoridou, M. (2025, April). *The AO-Cat ontology*. Zenodo. https://doi.org/10.5281/zenodo.15348438

May, K., Binding, C., & Tudhope, D. (2015). Barriers and opportunities for Linked Open Data use in archaeology and cultural heritage. *Archäologische Informationen*, *38*. https://doi.org/10.11588/ai.2015.1.26162

Marlet, O., Francart, T., Markhoff, B., & Rodier, X. (2019, June). *OpenArchaeo for usable semantic interoperability*. In Proceedings of the First International Workshop on Open Data and Ontologies for Cultural Heritage, co-located with the 31st International Conference on Advanced Information Systems Engineering (CAiSE 2019), Rome, Italy. CEUR Workshop Proceedings, Vol. 2375. http://ceur-ws.org/Vol-2375/

Nouvel, B., Granier, G., Boyer, C., & Anstett, E. (2021, November). *Terminology at the service of* *the characterisation of funerary practices: the Pactols Thesaurus*. HAL Open Science. https://shs.hal.science/halshs-03424453

Opiola, L., & Lamata Martinez, I. (2024, November). *D3.3 Final report on the EUreka3D services and resource hub: Design and implementation*. EUreka3D Project.

Quantin, M., Tournon, S., Chayani, M., Granier, X., & Laroche, F. (2025). Publishing and long-term archiving 3D data in humanities. In M. Ioannides, D. Baker, A. Agapiou, & P. Siegkas (Eds.), *3D research challenges in cultural heritage V: Paradata, metadata and data in digitisation* (pp. 75–88). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-78590-0_7

Quantin, M., Tournon, S., Grimaud, V., Laroche, F., & Granier, X. (2023). *Combining FAIR principles and long-term archival of 3D data*. ACM. https://doi.org/10.1145/3611314.3615913

Richards, J. D. (2023). Joined up thinking: Aggregating archaeological datasets at an international scale. *Internet Archaeology, 64*.

Salas Rossenbach, K. (2022, December). *D16.2 Final report on the ARIADNEplus pilots*. Zenodo. https://doi.org/10.5281/zenodo.7657027

Sanderson, R., Ciccarese, P., & Young, B. (2017). *Web Annotation Data Model*. W3C. Retrieved June 25, 2025, from https://www.w3.org/TR/annotation-model/

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data, 3*(1), 160018. https://doi.org/10.1038/sdata.2016.18