The AUTOMATA Metadata Scheme

"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."
(Richards, 2023, citing May, Binding & Tudhope, 2015)

This statement captures one of the key challenges in the digital heritage field: achieving real interoperability while avoiding the fragmentation of information.

Towards open and connected data

In line with the FAIR principles (Findable, Accessible, Interoperable, Reusable) and the broader goals of Open Science, the AUTOMATA project aims to make scientific and cultural data openly and transparently accessible.
For this reason, between May and June 2025, the Université Bordeaux Montaigne and the Archeovision team developed a metadata scheme that defines how information produced in AUTOMATA is structured, connected, and shared.
This framework ensures that data are traceable, interoperable, and preserved in the long term, fully aligned with the objectives of the European Collaborative Cloud for Cultural Heritage (ECCCH) and its flagship project ECHOES.

A shared structure for enriched digitisation

The metadata scheme provides a common language to describe and link all the information generated through the digitisation process, from 3D models to archaeometric data.
It builds on existing international standards such as Dublin Core, CIDOC-CRM, and the Europeana Data Model (EDM), ensuring that AUTOMATA data can be integrated within wider European research infrastructures.
Rather than introducing a new ontology, AUTOMATA promotes alignment with established frameworks like the AO-Cat ontology developed in ARIADNEplus, which supports searching by the powerful combination of *what*, *when*, and *where*.

Describing the data: What, Who, When, Where, and How

To make data easy to explore and reuse, the metadata scheme is structured around a set of fundamental questions:

- What – the *physical object*, such as an artefact or sample, identified by an inventory number and name.

- Who, When, Where – the *event*, recording who carried out the acquisition, and when and where it took place.

- How – the *technology*, describing how data were captured, including software, hardware, and acquisition processes.

Every acquisition is documented in detail, linking each 3D model to its corresponding object, instrument, and context. This approach guarantees traceability and scientific reproducibility.

From raw data to 3D models

The metadata scheme organises information around a digital object, a structured entity containing all data related to a single artefact, from the first capture to the final 3D model.
Raw data, such as photographs, are preserved unaltered and accompanied by all necessary capture parameters, calibration data, and paradata.
Each acquisition is spatially localised in 3D, linking the positions of robots and sensors to the resulting files. System logs and lighting metadata are automatically recorded to ensure accuracy and reproducibility.
Processed data, such as meshes, textures, and RIS3D (Referenced Information System in 3D), include mandatory metadata: the event (who, when, where), technology (how), file format, measurement units, and dimensions, along with optional details such as axis orientation or versioning.
By following shared practices like the Web Annotation Data Model (WADM) and IIIF 3D, AUTOMATA ensures that digital assets can be connected and annotated consistently across platforms.

Integrating metadata into the digitisation workflow

Metadata creation in AUTOMATA follows the entire digitisation sequence (before, during, and after acquisition).
Before digitisation, a *common event record* is created, including operator names, times, and locations, along with system setup and artefact information (inventory number, object type, discovery details).
During acquisition, metadata are generated automatically: every measurement receives a unique ID, and system logs record changes in parameters, lighting, and robot movement.
Computed data are documented with their processing methods and parameters.
After the sequence, all files and metadata are packaged into a digital container, assigned a DOI (Digital Object Identifier), and validated for completeness before submission to long-term repositories.
To assist with this process, AUTOMATA employs aLTAG3D, an open-source tool designed to verify, complete, and bundle metadata and data for preservation.

Organising and linking data

To support semantic consistency, the scheme adopts controlled vocabularies and thesauri, such as:

- the Getty Art & Architecture Thesaurus (AAT) for object types,
- GeoNames for geographic locations,
- Periodo for historical periods, and
- ORCID for identifying individuals.

These shared references, combined with Linked Open Data (LOD) and Uniform Resource Identifiers (URIs), make it possible to connect AUTOMATA datasets with other digital heritage repositories while keeping data distributed and accessible.

All technical details and diagrams are available in the full report D5.1 – Ontology and Metadata Scheme for Enriched Digitisation, accessible here.
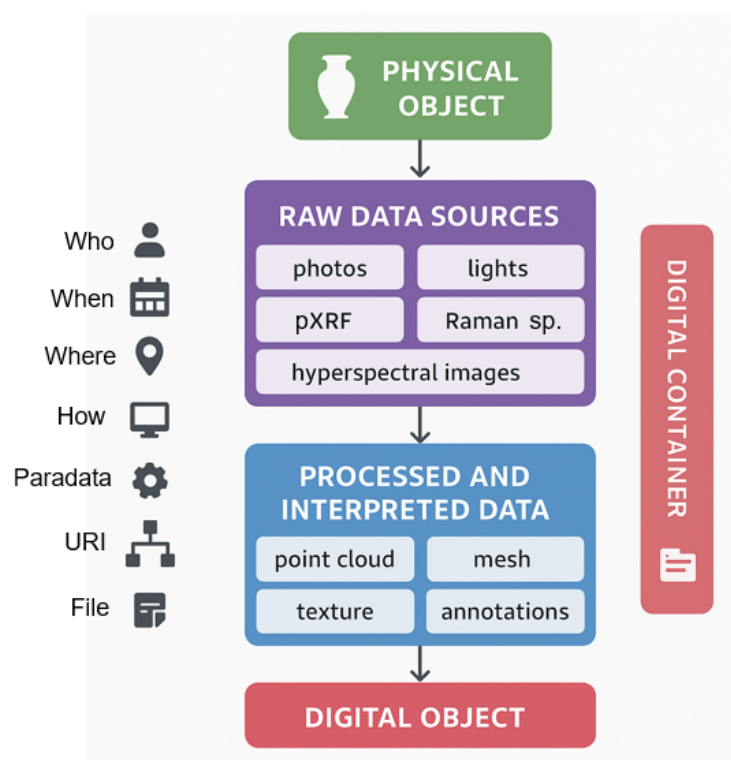

Fig. 1 - Each digital model is the result of a chain of actions that starts from the real object and moves through data collection and processing. Along the way, every step (who did it, when, with what tools, and how) is recorded.